OM protein - protein search, using sw model

Run on: December 29, 1999, 18:43:56 ; Search time 14.67 seconds
(without alignments)
589.327 Million cell updates/sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Title: US-09-077-173A-2
Perfect score: 1946
Sequence: 1 MASTESSLLRSLGLSPGPGS..........CRWAAPQDSSCSTPRADRF 365
Scoring table: BLOSUM62
Searched: 188963 seqs, 23686106 residues
Database :   A_Geneseq_36:*
Word size :   0
Number of hits that pass the threshold :   188963

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1940	99.7	365	1	W23606	Human P2Y4 recepto
2	928	47.7	375	1	W72457	Human P2O receptor
3	584	30.0	328	1	R91224	Mouse pancreas G-p
4	576	29.6	328	1	W09433	Human placenta G-p
5	571	29.3	328	1	R53752	Seven transmembran
6	510	26.2	339	1	W07617	Human G-protein th
7	510	26.2	339	1	W48733	Human R12 seven tr
8	510	26.2	344	1	W04247	Human G-protein co
9	478.5	24.6	344	1	W04733	Human 7-transmembr
10	449.5	23.1	370	1	W62597	Murine C140 recept
11	423	21.7	399	1	R66922	Murine C140 recept
12	423	21.7	399	1	W01954	Murine C140 recept
13	421	21.6	395	1	R66920	Murine C140 recept
14	418	21.5	395	1	W01952	Murine C140 recept
15	403	20.7	398	1	R66921	Human C140 recepto
16	403	20.7	398	1	R66921	Human C140 recepto
17	397.5	20.4	425	1	W01953	Human C140 recepto
18	397.5	20.4	425	1	R27240	Human thrombin rec
19	397.5	20.4	892	1	R60698	Fragment of the hu
20	397.5	20.4	425	1	W16314	Human thrombin rec
21	393.5	20.2	394	1	W51407	Human protease-act
22	393	20.2	397	1	W51408	Human protease-act
23	393	20.2	397	1	R66923	Human C140 recepto
24	387	19.9	372	1	R66922	Human C140 recepto
25	384	19.9	666	1	W01955	Rat delta opiate r
26	383	19.7	361	1	R76782	Leukocyte seven tl
27	383	19.7	361	1	W70504	Epstein Barr virus
28	380	19.7	333	1	R54080	Epstein Barr virus
29	380	19.5	400	1	W53623	Human T-cell recep
30	378.5	19.5	400	1	R63806	Human mu opioid re
31	378.5	19.5	372	1	R71966	Human mu opioid re
32	377	19.5	372	1	R76780	Human mu opiate re
33	377	19.4	356	1	R48629	Sequence of murine
34	377	19.4	372	1	R65188	Murine mu-subtype
35	377	19.4	372	1	R67670	Mouse delta opioid
36	376.5	19.4	398	1	W44938	Mouse delta opiate
37	376.5	19.3	398	1	R71964	Rat mu opioid rece
38	375.5	19.3	374	1	R76781	Rat mu opiate rece
38	375	19.3	398	1	W51406	Human protease-act
37	375.5	19.3	374	1	R76781	Human IP-10/Mig re
38	375	19.3	368	1	W54371	Human protease-act
39	372.5	19.1	371	1	R66503	Murine delta opiol
40	372.5	19.1	334	1	W22732	Human ATP receptor
41	371.5	19.1	342	1	R25835	Human platelet act
42	371.5	19.1	398	1	W44937	Mouse mu opiate re
43	366.5	18.8	334	1	W19854	Human purinergic r
44	366	18.8	415	1	W19780	Human G-protein ch
45	365.5	18.8	424	1	W30299	Human kappa opioid

ALIGNMENTS

RESULT 1
W23606

```
ID   W23606 standard; Protein; 365 AA.
AC   W23606;
DT   31-MAR-1998 (first entry)
DE   Human P2Y4 receptor polypeptide.
KW   Receptor; P2Y4; pyrimidine binding; uridine triphosphate; UTP.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FH
FT   Domain          35..61
FT                   /note= "membrane spanning domain I"
FT   Domain          72..95
FT                   /note= "membrane spanning domain II"
FT   Domain          113..131
FT                   /note= "membrane spanning domain III"
FT   Domain          151..174
FT                   /note= "membrane spanning domain IV"
FT   Domain          200..224
FT                   /note= "membrane spanning domain V"
FT   Modified_site   238
FT                   /note= "potential protein kinase C
FT                   (PKC) phosphorylation site"
FT   Modified_site   243
FT                   /note= "potential calmodulin-dependent protein kinase
FT                   and protein kinase C (PKC) phosphorylation site"
FT   Domain          247..271
FT                   /note= "membrane spanning domain VI"
FT   Domain          262
FT                   /note= "conserved His residue which is reported to
FT                   play a role in the P2Y2 receptor activation
FT                   by ATP and UTP"
FT   Modified_site   265
FT                   /note= "conserved Arg residue which is reported to
FT                   play a role in the P2Y2 receptor activation
FT                   by ATP and UTP"
FT   Domain          289
FT                   /note= "conserved Lys residue which is reported to
FT                   play a role in the P2Y2 receptor activation
FT                   by ATP and UTP"
FT   Modified_site   292
FT                   /note= "conserved Arg residue which is reported to
FT                   play a role in the P2Y2 receptor activation
FT                   by ATP and UTP"
FT   Domain          288..310
FT                   /note= "membrane spanning domain VII"
FT   Modified_site   345
FT                   /note= "potential protein kinase C (PKC)
FT                   phosphorylation site"
FT   Modified_site   359
FT                   /note= "potential protein kinase C (PKC)
FT                   phosphorylation site"
PN   WO9719170-A1.
PD   29-MAY-1997.
PF   21-NOV-1996; BE0123.
PR   21-NOV-1995; EP-870124.
PA   (EURO-) EUROSCREEN SA.
PI   Boeynaems J, Communi D, Parmentier M, Pirotton S;
DR   WPI; 97-402177/37.
DR   N-PSDB; T74321.
PT   Receptor having preference for pyrimidine over purine nucleotide(s)
PT   - especially uridine tri:phosphate, agonist and antagonists of which
PT   are useful in treatment of cystic fibrosis
```

```
CC   PS  Claim 1; Figure 1; 56pp; English.
CC   This sequence represents a novel human P2 receptor, P2Y4, which has a
CC   preference for pyrimidine binding, especially uridine triphosphate.
CC   This receptor could be used to screen for novel drugs which
CC   specifically bind to it. Transgenic animals could be used to determine
CC   the physiological effects of expressing varying levels of the receptor
CC   or to identify novel agonists or antagonists. The agonists and
CC   antagonists of human P2Y4 may be used, e.g., in treatment of cystic
CC   fibrosis.
SQ   Sequence  365 AA;
```

```
Query Match      99.7%;  Score 1940;  DB 1;  Length 365;
Best Local Similarity  100.0%;  Pred. No. 2.7e-208;
Matches  364;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy    1 MASTESSLLRSIGLSPGPGSSEVELDCWFDEDEKFILLPVSYAVVFVLGLNAPTLWLF 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MASTESSLLRSIGLSPGPGSSEVELDCWFDEDEKFILLPVSYAVVFVLGLNAPTLWLF 60

Qy   61 IFRLRPWDATATYMFHLALSDTLYVLSLPTLIYYAAHNHWPFGTEICKFVRFLFYWNLY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 IFRLRPWDATATYMFHLALSDTLYVLSLPTLIYYAAHNHWPFGTEICKFVRFLFYWNLY 120

Qy  121 CSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAWNLVVAGCLVPNLFYTTSNKG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAWNLVVAGCLVPNLFYTTSNKG 180

Qy  181 TTVLCHDTTRPEEEFDHYVHFSSAVWGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSR 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTVLCHDTTRPEEEFDHYVHFSSAVWGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSR 240

Qy  241 LRSLRTIAVVLTVFAVCFVPFHITRTIYYLARLLEADCRVLNIVNVVKYTRPLASANSC 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LRSLRTIAVVLTVFAVCFVPFHITRTIYYLARLLEADCRVLNIVNVVKYTRPLASANSC 300

Qy  301 LDPVLYLLITGDKYRRQLRQLCGGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCSTP 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 LDPVLYLLITGDKYRRQLRQLCGGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCSTP 360

Qy  361 RADR 364
        ||||
Db  361 RADR 364
```

```
RESULT 2

ID   R72457  standard; Protein; 375 AA.
AC   R72457;
DT   29-NOV-1995  (first entry)
DE   Human P2U-receptor.
KW   chronic bronchitis.
OS   Homo sapiens.
PN   WO9510538-A.
PD   20-APR-1995.
PF   04-OCT-1994;  U11260.
PR   15-OCT-1993;  US-138137.
PA   (UMOR ) UNIV MISSOURI.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Boucher RC; Erb LJ, Harden TK, Lustig KD, Parr CE;
PI   Sullivan DM, Turner JT, Weisman GA;
N-PSDB; Q08134.
DR   WPI; 95-169967/22.
PT   Disclosure; Page 34; 47pp; English.
CC   DNA encoding human P2U receptor and null cells expressing the
CC   receptors - for stimulating or inhibiting growth of cultures of
CC   mammalian cells, and for treating diseases of airway epithelial
CC   mucosa, e.g. asthma.
CC   The sequence is that of the human P2U receptor. The receptor may
CC   be used to treat diseases of epithelial mucosal surfaces by
CC   enhancing or inhibiting mucus prodn. It may be used to treat
CC   cystic fibrosis, asthma and chronic bronchitis.
SQ   Sequence  375 AA;
```

```
Query Match      47.7%;  Score 928;  DB 1;  Length 375;
Best Local Similarity   58.1%;  Pred. No. 2.2e-95;
Matches  182;  Conservative  41;  Mismatches  86;  Indels   4;  Gaps   3;
```

```
Qy   22 EVELDCWFDEDEKFILLPVSYAVVFVLGLNAPTLWLFIFRLRPWDATATYMFHLALSD 81
         |:| ||||||||||||||||||| || ||||| ||||:|||||||:|||||||||||:||
Db   20 ELGYRCRFNEDLKYVLLPVSYGVVCVLGLCINAVGLYIFLCLRKTWNASTTYMFHLAVSD 79

Qy   82 TLYVLSLPTLIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPL 141
         ||| ||||| :|||| |:||||   |||  ||||||||| ||: |||||||||:||| |
Db   80 ALYAASLPLLIVYYARGDHWPFSTVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPL 139

Qy  142 RALRWGRPRLAGLLCLAWNLVVAGCLVPNLFYTTSNKGTTVLCHDTTRPEEEFDHYVHFS 201
         |:||||| | | ||| ||: ||  |  |:|||:||:|  |:|||| :   |  |  | 
Db  140 RSLRWGRARYARRVAGAWVWLVLACQAPVLYFVTSARG-PLTCHDTISAPELFSRVAYS 198

Qy  202 SAVWGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSSRL--RSLRTIAVVLTVFAVCFV 259
         | :|||| |||  ::|||| ||||||  |: |  |:|||||   |||||||| |||:||:
Db  199 SVMLGLLFAVPFAVILVCYVLMARRLLKPAYGTSGGLPRAKRKSVRTIAVVLAVFALCFL 258

Qy  260 PFHITRTIYYLARLLEADCRVLNIVNVVKYTRPLASANSCLDPVLYLLITGDKYRRQLRQ 319
         |||:||::|| | || || :|:  :| | | || : |  ||||||||  | |:|||:|:
Db  259 PFHVTRTLYYSFRSIDLSCHTLNAINMAYKVTR-LASANSCLDPVLYFLAGQRLYRFARD 317

Qy  320 LCGGGKPQPRTAA 332
         | ||   |  |
Db  318 AKPPTGPSPAPPA 330
```

```
RESULT 3

ID   R91224  standard; Protein; 328 AA.
AC   R91224;
DT   26-AUG-1996  (first entry)
DE   Mouse pancreas G-protein coupled receptor protein.
KW   G-protein coupled receptor protein; G-PCRP; agonist; antagonist;
KW   cystic fibrosis; incontinence; diabetes; diagnosis; therapy.
OS   Mus sp.
PN   WO9605302-Al.
PD   22-FEB-1996.
PF   10-AUG-1995;  J01599.
PR   11-AUG-1994;  J01599.
PR   11-AUG-1994;  JP-189272.
PR   11-AUG-1994;  JP-189273.
PR   11-AUG-1994;  JP-189274.
PR   30-SEP-1994;  JP-236356.
PR   30-SEP-1994;  JP-236357.
PR   02-NOV-1994;  JP-270017.
PR   28-DEC-1994;  JP-326611.
PR   20-JAN-1995;  JP-007177.
PR   16-MAR-1995;  JP-057186.
PR   19-APR-1995;  JP-093989.
PA   (TAKE ) TAKEDA CHEM IND LTD.
PI   Fujii R, Fukusumi S, Hinuma S, Hosoya M, Ohgi K;
PI   Ohtaki T;
DR   WPI; 96-139698/14.
DR   N-PSDB; T18367.
PT   G-protein coupled receptor protein DNA and protein - also methods
PT   for isolating (ant)agonists for treatment of cystic fibrosis,
PT   incontinence and diabetes.
PS   Claim 6; Page 254-55; 360pp; English.
CC   A mouse pancreas beta-cell line MIN6-derived G-protein coupled
CC   receptor protein (G-PCRP) (R91224) was identified as the
CC   product of cDNA clone pMAH2-17 (T18367). The protein can be
CC   obtd. by expression of encoding cDNA in transformed host cells.
CC   It was classified into a subtype within the ATP receptor, P2
CC   purinoceptor. G-PCRPs (see also R91217-25 and R91227-33) can be
CC   used to screen agonists and antagonists that modulate G-PCRP
CC   activity, to raise antibodies and to develop assay systems.
SQ   Sequence  328 AA;
```

```
Query Match         30.0%;   Score 584;  DB 1;  Length 328;
Best Local Similarity  41.6%;  Pred. No. 4.5e-57;
Matches 134;  Conservative 40;  Mismatches 132;  Indels 16;  Gaps 6;

Qy    5  ESSLLRSIGLSPGPGGSSEVELDCWFEDEDFKFILLPVSIAVVFVLGLGLNAPTLWLFIFRL  64
         :||:|   |  |||||||||| |::| || |:|:||| |:|||||||::||||||||||||
Db    4  DNGTIQAPGLPP-----------TTCVYREDFKRLLLTPVSVSVLVVGLIPLNICVIAQICASR  55

Qy   65  RPWDATAYMFHLALSDTLIYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVL   124
         ::|:|   ||  ||||||| |||| | || :| :|::|||||| ||||||||||| ||:|
Db   56  RTLTRSAVYTLNIALADLMYACSLPLLIIYYARGDHWPFGDLACRFVRFLFYANLHGSIL   115

Qy  125  FLTCISVHRYLGTCHPLRALRW---GRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGT   181
         |||||||||||| |||||||||   |||||||| ||||||||||||||||||||||::|
Db  116  FLTCISFQRYLGICHPLAS--WHKRGGRRAAWVCGVVWLAVTAQCLPTAVFAATGIQRN   173

Qy  182  TVLCHDTTRPEEFDHVHFSSAVMGLLFGVPCLVTLVCYGLMARRL-YQPLP-GSAQSSS   239
         |||| ||||||||||||||||||||||||||||||:||| ||||  | |:| | |
Db  174  RRVCYDLSPPILSTRYLPYGMALTVIGFLLPFIALLACYCRMARRLCRQDGPAGPVAQER   233

Qy  240  RLRSLRTIAVVLITVFAVCGVPEHITRTIYYLARLLE-ADCRVLNIVNVVKVTRPLASAN   298
         |||   |:  :|| |  || ||| :| :|||||| || |  |||| ||| || |||
Db  234  RSKARMAVVVAAVFAISFLPFHITKTAYLAVRSTPGVSGVCPVLETFAAAYKGTRPFASVN   293

Qy  299  SCLDPVLILTGDKYRQLRQL  320
         ||:|:|    |||||||||
Db  294  SVLDPILFYFTQQKFRQPHDL  315
```

```
RESULT   4
R91225

ID   R91225 standard; Protein; 328 AA.
AC   R91225;
DT   26-AUG-1996   (first entry)
DE   Human placenta G-protein coupled receptor protein.
KW   G-protein coupled receptor protein; G-PCRP; agonist; antagonist;
KW   cystic fibrosis; incontinence; diabetes; diagnosis; therapy.
OS   Homo sapiens.
PN   WO9605302-A1.
PD   22-FEB-1996;
PF   10-AUG-1995;  J01599.
PR   11-AUG-1994;  J01599.
PR   11-AUG-1994;  JP-189272.
PR   11-AUG-1994;  JP-189273.
PR   11-AUG-1994;  JP-189274.
PR   30-SEP-1994;  JP-236356.
PR   30-SEP-1994;  JP-236357.
PR   02-NOV-1994;  JP-270017.
PR   28-DEC-1994;  JP-326611.
PR   20-JAN-1995;  JP-007177.
PR   16-MAR-1995;  JP-057186.
PR   19-APR-1995;  JP-093989.
PA   (TAKE ) TAKEDA CHEM IND LTD.
PI   Fujii R, Fukusumi S, Hinuma S, Hosoya M, Ohgi K;
PI   Ohtaki T;
DR   WPI; 96-139698/14.
DR   N-PSDB; T18368.
PT   G-protein coupled receptor protein DNA and protein - also methods
PT   for isolating (ant)agonists for treatment of cystic fibrosis,
PT   incontinence and diabetes
PS   Claim 6; Page 267-68; 360pp; English.
CC   A novel human placenta-derived G-protein coupled receptor protein
CC   (G-PCRP) (R91225) was identified as the product of cDNA clone
CC   phAH2-17 (T18368). The protein can be obtd. by expression of the
CC   cDNA clone in transformed host cells. It was classified as a
CC   purinoceptor. G-PCRPs (see also R91217-24 and R91227-33) can be
CC   used to screen agonists and antagonists that modulate G-PCRP
CC   activity, to raise antibodies and to develop assay systems.
SQ   Sequence  328 AA;

Query Match         29.6%;   Score 576;  DB 1;  Length 328;
Best Local Similarity  41.8%;  Pred. No. 3.5e-56;
```

```
Matches 133;  Conservative 38;  Mismatches 129;  Indels 18;  Gaps 6;

Qy   10  RSLGLSPGPGGSSEVELDCWFEDEDFKFILLPVSIAVVFVLGLGLNAPTLWLFIFRLRPWDA  69
         :||||  |  ||||||||||| | | || |:|:|| |:|||||||::|||||||||||
Db    9  QALGLPP-----------TTCVYRENFKQLLLPPVSYAVVFVLGLNAPTLLAAGLPLNICVIQICTSRRALTR  60

Qy   70  TATYMFHLALSDTLIYVLSLPTLIIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCI  129
         :||:|| ||||||| |||| | || :|:|:||||||| |||||||||||| ||:||||||
Db   61  TAVYTLNIALADLIYACSLPLLIIYYAQGDHWPFGDFACRLVRFLFYANLHGSILFLTCI   120

Qy  130  SVHRYLGICHPLRALRW---GRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCH  186
         |||||| |||||||||    |||||||| |||||||||||||||||||||||||: |
Db  121  SFQRYLGICHPLAP--WHKRGGRRAAWLVCVTVWLAVTTQCLPTAIFAATGIQRNRTVCY  178

Qy  187  DTTRPEEFDHVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQ---PLPGSAQSSSRLRS  243
         ||||||||||||||||||||||||||||||||| |||| |    | | |||||||||
Db  179  DLSPPALATHYMPYGMALTVIGFLLPFAALLACYCLLACRLCRQDGPAEPVAQ-ERGKA  237

Qy  244  LRTIAVVLITVFAVCGVPEFHITRTIIY-LARLLEADCRVLNIVNVVKVTRPLASANSCLD  302
         |||| ||| || || ||| :|| ||||  || | |||||| || || ||| |||
Db  238  ARMAVVVAAAFAISFLPFHIKTKTAYLAVLAVGSTPGVPCTVLEAFAAAYKGTRPFASANSVLD  297

Qy  303  PVLILTGDKYRQLRQLRQL  320
         ||:|:|    ||||||| |
Db  298  PILFYFTQKKFRRRHEL  315
```

```
RESULT   5
W09433

ID   W09433 standard; Protein; 328 AA.
AC   W09433;
DT   01-SEP-1997   (first entry)
DE   Human placenta purinergic P-2u receptor, PNR.
KW   PNR; purinergic P-2u receptor; placental; inflammation; diagnosis;
KW   carcinoma; neoplasia; cancer; cystic fibrosis; hypertension;
KW   high blood pressure; infection.
OS   Homo sapiens.
PN   WO9638558-A2.
PD   05-DEC-1996;
PF   03-JUN-1996;  U08481.
PR   02-JUN-1995;  US-459046.
PA   (INCY-) INCYTE PHARM INC.
PI   Au-Young J, Coleman R, Stuart SG;
DR   WPI; 97-042714/04.
DR   N-PSDB; T47730.
PT   New isolated human placenta purinergic P-2u receptor (PNR) gene -
PT   used to develop prods. for the diagnosis and treatment of conditions
PT   associated with altered expression of the receptor e.g. inflammation
PS   Claim 1; Page 27-29; 36pp; English.
CC   W09433 shows a human placenta-derived purinergic P-2u receptor
CC   designated PNR. P-2u is specifically expressed in cells active in
CC   immunity. An assay for up-regulated expression of PNR can accelerate
CC   diagnosis and proper treatment of conditions caused by abnormal signal
CC   transduction due to systemic and local infections, traumatic and
CC   other tissue damage, hereditary or environmental diseases associated
CC   with hypertension, carcinomas, cystic fibrosis and other physiological
CC   or pathological problems.
SQ   Sequence  328 AA;

Query Match         29.3%;   Score 571;  DB 1;  Length 328;
Best Local Similarity  41.8%;  Pred. No. 1.3e-55;
Matches 133;  Conservative 36;  Mismatches 131;  Indels 18;  Gaps 6;

Qy   10  RSIGLSPGPGGSSEVELDCWFEDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDA  69
         ::||||  |  ||||||||||| | | |::| |:|:|| |:|||||||::||||||||||||
Db    9  QALGLPP-----------TTCVYRENFKQLLLPPVSAVLAPALPLNICVITQICTSRRALTR  60

Qy   70  TATYMFHLALSDTLIYVLSLPTLIIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCI  129
         |:| |:|| ||||||| |||| | || :| :|::|||||| ||||||||||| ||:||||||
Db   61  TAVYTLNIALPDLLYACSLPLLIIYYAQGDHWPFGDFACRLVRFLFYANLHGRILFLTCI   120

Qy  130  SVHRYLGICHPLRALRW---GRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCH  186
         |||||| |||||||||    |||||||| |||||||||||||||||||||||||: |
```

```
Db  121 SFQRYLGICHPLAP--WHRRGRRAAWLVCVAVWLAVTTQCLPTAIFAATGIQRNRTVCY 178
        |||||||||      :|        | :| | ||||| :|  |::|  |  : ||:
Qy  187 DTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCGIMARRLYQ---PLPGSAQSSSRLRS 243

Db  179 DLSPPALATHYMPYGMALTVIGFLLPFAALLACYCLLACRLCRQDGPAEPVAQ-ERRGKA 237
Qy  244 LRTIAVVLIVFAVGVPFHTRTIIYYLARLLE-ADCRVLNIVNVVYKVTRPLASANSCLD 302

Db  238 ARMAVVAAVFGISFLPFHITKTAYLAVRSTPGVPCTVLEAFAAAYKGTRPFAASANSVLD 297
Qy  303 PVLYLLTGDKYRQLRQL 320

Db  298 PILFYFTQKFRRPHEL 315
```

RESULT 6
R53752

```
ID   R53752 standard; Protein; 339 AA.
AC   R53752;
DT   07-FEB-1995 (first entry)
DE   Seven transmembrane receptor (R12).
OS   Homo sapiens.
KW   Primer; seven transmembrane receptor; receptor; amplification.
FH   Key              Location/Qualifiers
FT   domain           33..57
FT                    /label= Transmembrane domain.
FT   domain           68..90
FT                    /label= Transmembrane domain.
FT   domain           106..127
FT                    /label= Transmembrane domain.
FT   domain           145..168
FT                    /label= Transmembrane domain.
FT   domain           193..217
FT                    /label= Transmembrane domain.
FT   domain           233..251
FT                    /label= Transmembrane domain.
FT   domain           290..312
FT                    /label= Transmembrane domain.
PN   WO9412635-A.
PD   09-JUN-1994.
PF   17-NOV-1993; U11153.
PR   17-NOV-1992; US-977452.
PA   (ICOS-) ICOS CORP.
PI   Godiska R, Gray PW, Schweickart VL;
DR   WPI; 94-200264/24.
DR   N-PSDB; Q66178.
PT   DNA encoding seven trans:membrane receptors - used to develop
PT   prods. for use as therapeutic or diagnostic agents for conditions
PT   involving the receptors.
PS   Example 10; Page 81-82; 100pp; English.
PT   DNA encoding seven transmembrane coding sequence
CC   partial sequence of the R20 seven transmembrane receptor.
CC   which was later used as a probe for isolating the R20 genomic clone
CC   (Q66176) from a human placenta DNA library. During the isolation
CC   of the R20 gene, two weakly hybridising sequences were identified
CC   which had significant homology to other seven transmembrane
CC   receptors. The probe was used to screen a human genomic foetal
CC   liver DNA library, and while the R20 gene could not be identified in
CC   this library, several weakly hybridising clones were plaque
CC   purified, subcloned and sequenced. The two clones were designated
CC   R2 (Q66177) and R12. This is the sequence encoded by the R12 clone.
SQ   Sequence 339 AA;
```

```
Query Match      26.2%;  Score 510;  DB 1;  Length 339;
Best Local Similarity   37.1%;  Pred. No. 8.5e-49;
Matches 119;  Conservative 52;  Mismatches 136;  Indels 14;  Gaps 7;

Qy   27 CWFEDEFKFILLPVSYAVVFVGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVL 86
Db   23 CGQETPLENMLFASFYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVL 82
```

```
Qy   87 SLPTLIYYAAHNHWPFGTEICKFVRFLFYWNLCSVLFLTCISVHRYLGICHPLRALRW 146
Db   83 VLPTRLIVYHFSGNHWPFGEIACKRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKL 142

Qy  147 GRPRLAGLLCLAWVLVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPEEFDHYVHFSSAVMG 206
Db  143 RRPLYAHLACAFLWVVAVAMAPLLVSPQTVQNTHTVVCLQLYR-EKASHHALVSLAV-- 200

Qy  207 LLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVGVPFHITRT 266
Db  201 -AFTPPFITTVTCLLIIIRSLRQGL--RVEKRLRTKAVRMAIVLAIFLCVFVPHVNRS 256

Qy  267 IYYL-ARLLEADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTGDKYRQL-RQLCG-- 323
Db  257 VYVLHYRSHGAyCApCAFQRILALANRITSCLISLNGALDPIMYFFVAEKFRHALCNLLCGKR 316

Qy  323 --GGKP--QPRTAASSLALVS 339
Db  317 LKGPPPSFEGKTNESSLSAKS 337
```

RESULT 7
W07617

```
ID   W07617 standard; Protein; 339 AA.
AC   W07617;
DT   28-FEB-1997 (first entry)
DE   Human G-protein thrombin-like receptor.
OS   Homo sapiens.
KW   G-protein; thrombin; receptor; diagnosis; therapy; agonist;
KW   antagonist; haemophilia; wound healing; restenosis; angina;
KW   inflammation.
PN   WO9639438-A1.
PD   12-DEC-1996.
PF   06-JUN-1995; U07180.
PR   06-JUN-1995; WO-U07180.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Gocayne JD, Li Y, Ruben SM;
DR   WPI; 97-043073/04.
DR   N-PSDB; T44092.
PT   Human G-protein thrombin receptor, HIBEB69 - useful to identify
PT   (antagonists, for treatment of angina, restenosis, wound healing
PT   etc.
Claim 1; Fig 1A-D; 58pp; English.
CC   A novel human 7-transmembrane receptor (W07617) has been putatively
CC   identified as a G-protein thrombin receptor. Its amino acid
CC   sequence was deduced from a cDNA clone (T44092) discovered in a
CC   human infant brain cDNA library.  Recombinant receptor can be
CC   produced in host cells (e.g. E. coli, COS-7, SF9) and used to
CC   raise antibodies or to screen for (antagonist cpds. Agonists
CC   can be used to treat e.g. haemophilia or to promote wound
CC   healing; antagonists can be used to treat inflammation, angina,
CC   restenosis, stoke, etc.
SQ   Sequence 339 AA;
```

```
Query Match      26.2%;  Score 510;  DB 1;  Length 339;
Best Local Similarity   37.1%;  Pred. No. 8.5e-49;
Matches 119;  Conservative 52;  Mismatches 136;  Indels 14;  Gaps 7;

Qy   27 CWFEDEFKFILLPVSYAVVFVGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVL 86
Db   23 CGQETPLENMLFASFYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVL 82

Qy   87 SLPTLIYYAAHNHWPFGTEICKFVRFLFYWNLCSVLFLTCISVHRYLGICHPLRALRW 146
Db   83 VLPTRLIVYHFSGNHWPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKL 142

Qy  147 GRPRLAGLLCLAWVLVVAGCLVPNLFYYTTSNKGTTVLCHDTTRPEEFDHYVHFSSAVMG 206
Db  143 RRPLYAHLACAFLWVVAVAMAPLLVSPQTVQNHTVVCLQLYR-EKASHHALVSLAV-- 200

Qy  207 LLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVGVPFHITRT 266
```

```
Db  200 -AFTPFFITIVTCYLLLIIRSLRQGL--RVEKRLKTKAVRMIAIVLAIFLVCFVPYHVNRS 256
         :||:| :|||::          :|  ::|| |:: | |  |:| ||||:||||||:|
Qy  267 IIYL-ARLLEADCRVLNIVIVNVYVKYRPLASANSCLDPVLYLLTGDKYRRQL-RQLCG-- 323
         :|| : :     ::|  |:    |:     :::|| |  :|||  :|   |  :||
Db  257 VYVLHYRSHGASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKR  316

Qy  323 --GGKP--QPRTAASSIALVS 339
           ||||    |    ::| |
Db  317 LKGPPPSFEGKINESSLSAKS 337
```

RESULT 8

```
ID   W48733   standard; Protein; 339 AA.
AC   W48733;
DT   25-SEP-1998  (first entry)
DE   Human R12 seven transmembrane (7TM) receptor.
KW   V28; placenta; seven transmembrane receptor; 7TM; signal transduction;
KW   Immunology; inflammation; R12.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Domain          33..57
FT                   /note= "Transmembrane domain 1"
FT   Domain          68..90
FT                   /note= "Transmembrane domain 2"
FT   Domain          106..127
FT                   /note= "Transmembrane domain 3"
FT   Domain          145..168
FT                   /note= "Transmembrane domain 4"
FT   Domain          193..217
FT                   /note= "Transmembrane domain 5"
FT   Domain          233..251
FT                   /note= "Transmembrane domain 6"
FT   Domain          290..312
FT                   /note= "Transmembrane domain 7"
PN   US5775980-A.
PD   02-JUN-1998;
PF   17-NOV-1993; 153848.
PR   17-NOV-1992; US-977452.
PA   (ICOS-) ICOS CORP.
PI   Godiska R, Gray PW, Schweickart VL;
DR   WPI; 98-332132/29.
DR   N-PSDB; V18356.
PT   DNA encoding V28 seven transmembrane receptor polypeptide - useful
PA   for producing recombinant polypeptide and anti-V28 antibodies, and
PR   for use in screening assays for V28 agonists and antagonists
PS   Example 10; Columns 87-90; 56pp; English.
CC   The present sequence represents the human R12 protein which is encoded
CC   by the full length genomic R12 clone (V18356) which was isolated from
CC   a human genomic fetal liver DNA library. The invention claims for the
CC   full length V28 genomic DNA sequence (V18343) isolated from a human
CC   placenta genomic library. The V28 (W48722) and R12 proteins are seven
CC   transmembrane (7TM) receptors which are probably involved in signal
CC   transduction. The invention also claims that cells transformed with V28
CC   DNA can be used to produce the recombinant polypeptide, to produce
CC   anti-V28 antibodies or in screening assays for V28 agonists or
CC   antagonists. The antibodies, agonists and antagonists could then be
CC   used to modulate V28 receptor-ligand binding, for e.g. in Immunological
CC   and/or inflammatory events in vivo.
SQ   Sequence 339 AA;
```

Query Match 26.2%; Score 510; DB 1; Length 339;
Best Local Similarity 37.1%; Pred. No. 8.5e-49;
Matches 119; Conservative 52; Mismatches 136; Indels 14; Gaps 7;

```
Qy  27 CWFEDEDFKFILLPVSYAVFVLGLGLNAPTLWLFIFRLRFWDATATYMFHLALSDTLYVL  86
        ||||  |||:|| |::||:|| ||||||| |||:||  |:|||:|||||  :||  ||||
Db  23 CGQEFPLENMLFASFYLLDFILALVGNTLLAWLFIRDHKSCTPANVFLMHLAVADLSCVL  82

Qy  87 SLPTLIIYYLYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRW 146
        .|| |||:||  :|| |::|    |:  ::|||::|:|:||||||:  ||  |  ||:  |
Db  83 VLPTRLVYHFSGNHWPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIAHPVKSLKT 142
```

```
Qy  147 GRPRLAGLLCLAVWLVVAGCLVPNLFVTTSNKGTTVLCHDTTRPEEFDHYVHFSSAVMG 206
         |||| |:|  ||::||| || :| |||:|||:| ||||||:|||||||:|:||||||:|
Db  143 RRPLYAHLACAFLMVVVAVAMAPLLVSPQTVQTNHTVCLQLYR-EKASHALVSLAV-- 200

Qy  207 LLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHTRTI 266
         : |  | :|:|||||||||| |: :  |::|||||:||||:|:|| |||||||:|  |
Db  200 -AFTPFFITIVTCYLLLIIRSLRQGL--RVEKRLKTKAVRMIAIVLAIFLVCFVPYHVNRS 256

Qy  267 IIYL-ARLLEADCRVLNIVIVNVYVKYRPLASANSCLDPVLYLLTGDKYRRQL-RQLCG-- 323
         :||:| :|||::          :|  ::|| |:: | |  |:| ||||:||||||:|
Db  257 VYVLHYRSHGASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKR  316

Qy  323 --GGKP--QPRTAASSIALVS 339
           ||||    |    ::| |
Db  317 LKGPPPSFEGKINESSLSAKS 337
```

RESULT 9

```
ID   W04247   standard; Protein; 344 AA.
AC   W04247;
DT   13-DEC-1996  (first entry)
DE   Human G-protein coupled receptor GPR4.
KW   G-protein coupled receptor; GPR4; signal transduction; agonist;
KW   antagonist; cell proliferation; cancer; tumour; asthma; allergy;
KW   diagnosis.
OS   Homo sapiens.
PN   WO9630406-A1.
PD   03-OCT-1996;
PF   30-MAR-1995; U04079.
PR   30-MAR-1995; WO-U04079.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Sutton GG;
PI   Built CJ, Cao L, Gentz R, Li Y, Ni J, Rosen CA;
DR   WPI; 96-455278/45.
DR   N-PSDB; T33905.
PT   Novel human G-protein coupled receptor - for
PT   diagnosing diseases, and identifying (ant)agonists for asthma and
PT   allergy treatment
PS   Claim 1; Page 56-57; 87pp; English.
CC   Novel human G-protein coupled receptor GPR4 (W04247) was identified
CC   as the product of a cDNA clone (T33905) isolated from a human 12-wk
CC   embryo. The protein is also expressed in the human heart, spleen
CC   and leukocytes. It shows 82% identity over a 291-amino acid
CC   stretch with a chick orphan G-protein coupled receptor. Potential
CC   ligands include thrombin, chemokine and platelet activating
CC   factor. Recombinant GPR4, GPR1, GPR2 and GPR3 (see also W04244-46)
CC   can be expressed in e.g. E. coli, COS or insect cell hosts for use in
CC   identifying (antagonist cpds. Agonists may be used to treat asthma,
CC   Parkinson's disease, hypotension, osteoporosis etc., and antagonists
CC   to treat ulcers, asthma, allergies, etc.
SQ   Sequence 344 AA;
```

Query Match 24.6%; Score 478.5; DB 1; Length 344;
Best Local Similarity 33.4%; Pred. No. 2.8e-45;
Matches 99; Conservative 66; Mismatches 126; Indels 5; Gaps 4;

```
Qy   27 CWFEDEDFKFILLPVSYAVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDLLFVF  68
          | :::    : | |:  ::|||||:|| |::: :|| |:|:  ::||:|:|:|:||| |
Db    9 CFYNDSPKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSDLLFVL  68

Qy   87 SLPTLIIYYLYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRW 146
          :||  |:|   ::| |||| :|| | :::|::|:|||:|||||||||:| |||  :   :
Db   69 TLPFRIFYFTTRN-WPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIYPFKSKTL 127

Qy  147 GRPRLAGLLCLAVWLVVAGCLVPNLFVTTSNKGTTV--LCHDTTRPEEFDHYVHFSSAV 204
         |  | |   | ||: |: |   |  | :     ::   | |||:|||| ||||||||||
Db  128 RTKRNAKIVCTGVWLTVIGGSAPAVFEVQSTHSQGNNASEACFENFPEATWKITLSRIVIF 187

Qy  205 MGLL--FGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHI 263
         :|:|   |||| :|:|| |:|||||||||||||||||||||||||||||||||||||||||
Db  205 VLPTLVIYHFSGNHWPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIAHPVKSLKT 263
```

RESULT 10

```
Db   188  IEIVGFFIPLILNVTCSSMVLKILTKPVTLSRSKINKIKIVLKMIFVHLIIFCFGFVPYNI  247
Qy   264  TRTIYYLARLLE-ADCRVLNIVNNVYKVTRPLASANSCLDPVLYLLTGDKYRRQLR       318
Db   248  NLIIYSLVRTQTEVNCSVVAAVRIMYPITLCIAVSNCCPDIVYYFTSDTIQNSIK        303
```

ID   W62597 standard; Protein; 370 AA.
AC   W62597;
DT   07-OCT-1998 (first entry)
DE   Human 7-transmembrane receptor;
KW   G-protein coupled receptor; HE8CS41; treatment; diagnosis; infection;
KW   HIV-1; HIV-2; cancer; screening; human; 7-transmembrane receptor;
KW   Parkinson's disease; vaccine; anorexia; bulimia; hypertension; asthma;
KW   urinary retention; osteoporosis; ulcer; allergy; neurological disorder.
OS   Homo sapiens.
PN   EP-853116-A2.
PD   15-JUL-1998;
PF   08-JAN-1998; 300130.
PR   10-JAN-1997; US-781250.
PA   (SMK) SMITHKLINE BEECHAM CORP.
PI   Bergsma DJ, Mao JY, Sathe G, Van HORN S;
DR   WPI: 98-364651/32.
DR   N-PSDB; V38565.
CC   Human G-protein coupled receptor, HE8CS41 - useful e.g. in treatment
CC   and diagnosis of infections e.g. by HIV-1, HIV-2, cancers etc. and
CC   screening of antagonistic or agonistic compounds
CC   Claim 15; Fig 1; 20pp; English.
PS   Fig 1; 20pp; English.
PT   Human G-protein coupled receptor, HE8CS41. HE8CS41 is a
PT   human G-protein coupled receptor and can be used in vaccines to protect
PT   mammals from infections e.g. bacterial, fungal, protozoan and viral
PT   infections (especially infections caused by HIV-1 or HIV-2), cancers and
PT   Parkinson's disease. They can also be used in the treatment of anorexia,
PT   bulimia, hypertension, asthma, urinary retention, osteoporosis, ulcers,
PT   allergy and neurological disorders. The HE8CS41 polypeptides can be used
PT   to produce antibodies which are used to treat such infections, isolate
CC   or identify clones expressing the polypeptide or to purify the
CC   polypeptide. The polypeptides can also be used to screen for compounds
CC   binding to, activating or inhibiting activation of HE8CS41. They can be
CC   used to diagnose diseases and susceptibility to diseases related to
CC   expression or activity of HE8CS41.
SQ   Sequence 370 AA;

Query Match   23.1%; Score 449.5; DB 1; Length 370;
Best Local Similarity 33.4%; Pred. No. 5.4e-42;
Matches 102; Conservative 67; Mismatches 131; Indels 5; Gaps 5;

```
Qy    14  LSPGPGSSEVELDCWFDEDFKIFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATAYY   73
              |:| . |:| |:| |:| |||:| |:||| : |:|:|||::
Db    18  LRPLGNATANNTCIVDDSFKYNLNGAVYSVVFIILGLTINSVSLFVFCFRMKRSETAIF    77

Qy    74  MFHLALSDTLYVLSLPFLIYYYAAHHNWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHR  133
              :|||.|||||:|:|:|||  | |:||| |:|::|| |:|||| :|:|:||
Db    78  ITNLAVSDLLFVCTLPFKI-FYNFNRHWPFGDTLCKISGTAFLTNIYGSMLFLTCISVDR  136

Qy   134  YLGICHPLRALRNGRPRLAGLLCLAVW-LVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPE  192
              : |:| .|| | |:|| | |:| |||:||| | |:| |:||
Db   137  FLAIVYPPRSRIRTRNSAIVCAGVWILVLSGGISASLFSTINVNNATTT-CFEGFSKR    195

Qy   193  EFDHYVHFSSAVMGLL-FGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSRTIAVVL   251
              |:| . | :| . |:| |:| |:| |:|:|| |:|
Db   196  VWKTYLSKITIFIEVGFIIPLLNVSGCSSVVLRTLRKPATLSQIGTNKKKVLKMITVHM   255

Qy   252  TYFAVCFVPFHITRTIYYLARLLE-ADCRVLNIVNNVYKVTRPLASANSCLDPVLYLLTG  310
              ||:| |:|| |:| |:| : |:|||| |:|:|
Db   256  AVFVCCFVPYNSVLFLYALVRSQAITNCFLERFAKIMPITLCLATLNCCFDPIYYFTL    315

Qy   311  DKYRR  315
              : :||
Db   316  ESFQK  320
```

RESULT 11

ID   R66922 standard; Protein; 399 AA.
AC   R66922;
DT   22-AUG-1995 (first entry)
DE   Murine C140 receptor deduced from cDNA.
KW   G-protein coupled receptor; G-protein; C140 receptor.
OS   Mus musculus.
PN   WO9503318-A.
PD   02-FEB-1995;
PN   26-JUL-1994; U08536.
PF   26-JUL-1993; US-097938.
PA   (CORT-) COR THERAPEUTICS.
PI   Scarborough RM, Sundelin J;
     WPI: 95-075182/10.
     N-PSDB; Q84559.
CC   New DNA encoding recombinant C140 receptor - and novel agonists
CC   and antagonists and specific antibodies with therapeutic and
CC   diagnostic applications.
PS   Example; Fig 10; 57pp; English.
PT   A cDNA library from a mouse stomach was constructed in lambda gt10
CC   and screened with a probe encompassing the C140 genomic clone (see
CC   Q84557). A single phage clone was isolated and cut with EcoRI. The
CC   insert was cloned in pBluescript and pSG5 and sequenced. The
CC   complete nt sequence and deduced AA sequence is given in Q84559 &
CC   R66992. RACE resulted in the addition of only 27 bps to the 5'
CC   end. The 5' RACE resulted in the addition of only 27 bps to the 5'
CC   end of the apparent coding region differs from the 5'
CC   end of the ORF of genomic DNA; it is believed that the 5' end of
CC   the cDNA sequence is correct.
SQ   Sequence 399 AA;

Query Match   21.7%; Score 423; DB 1; Length 399;
Best Local Similarity 31.8%; Pred. No. 5.3e-39;
Matches 107; Conservative 55; Mismatches 141; Indels 34; Gaps 10;

```
Qy    36  ILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPFLIYY    95
              ||:| | |:| | |:| |:||| | |:|:|:||| |:|
Db    78  VFLPVYIVFVIGLPSNGMALWIFLFRTKKHPAVIMANLALADLLSVIWFPLKISYH     137

Qy    96  AAHHNWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPR---L  151
              |:|||| |:| |:||| :| |:| |:| | | |:||
Db   138  LHGNNWVYGEALCKVLIGFFYGNNYCSILFMTCLSVQRYWVIVNPM-----GHPRKKANI 192

Qy   152  AGLLCLAVWLVVAGCLVP------NLFFVTTSNKGTTVLCHDTTRPEEF---DHVHFSSA  203
              | |:| |:| |:| | |:|||| |:| |:|
Db   193  AVGVSLAIWLLILFLYVTIPLIYVMKQTIYIPALNITT---CHDVL-PEEVLVGDMFNFLSL 248

Qy   204  VMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSRTIAVVLLITVFAVCFVPFHI 263
              |:| |:| |:||| |:| |:| |:|:|||
Db   249  AIG-VFLPALLTASAYVLMIKILRSSAMDEHSEKRQRAIRLIITVLAMYFICFAPSNL   307

Qy   264  TRTIYYLARLLEADCRVLNIVNNVYKVTRPLASANSCLDPVLYLLTGDKYRRQLR--LC   321
              || : | | |:| |:| |:||||
Db   308  LLIVHYFLIKTQRQSHVY----ALIYLVALCLSTLNSCIDPFVYYFVSKDFRDHARNALLC 363

Qy   322  GGGRKPQPRFAASSLALVSLPEDSSCRWAATPQDSSCS  358
              | |:|| ||
Db   364  RSVRTVNRMQ------ISLSSNKFSRKSGSYSSSSTS   394
```

RESULT 12

ID   W01954 standard; Protein; 399 AA.
AC   W01954;
DT   02-APR-1997 (first entry)
DE   Murine C140 receptor.
KW   C140 receptor; G-protein linked; coupled; seven pass; agonist;
KW   antagonist; hypertension; hypotension; blood pressure.
OS   Mus sp.
FH   Key   Location/Qualifiers

```
FT   peptide         1..31
FT                   /note= "the signal peptide differs from that
FT                   encoded by a genomic DNA sequence for
FT                   this receptor (see WO1952), the signal
FT                   sequence given here is believed to be
FT                   the correct sequence"
FT   protein         32..399
FT                   /note= "mature protein"
PN   WO9623225-A1.
PD   01-AUG-1996.
PF   25-JAN-1996; U01179.
PR   25-JAN-1995; US-390301.
PA   (CORT-) COR THERAPEUTICS INC.
PI   Scarborough RM, Sundelin J;
DR   N-PSDB; T32038.
WPI; 96-362813/36.
PS   Example 4; Fig 10A-B; 60pp; English.
CC   WO1954 represents the murine C140 receptor.
CC   W01954 represents the murine C140 receptor.
PT   Vector for expression C140 cell surface receptor in host cell -
PT   useful to identify C140 agonist and antagonists, which are
PT   antihypertensives and elevators of blood pressure, respectively
CC   DNA encoding C140R
CC   may be engineered so as to allow the recombinant expression of C140R in
CC   a suitable host cell, i.e. by removing the native expression-control
CC   sequences and replacing them with control sequences operable in the host.
CC   Such a recombinant receptor can be expressed on the surface of oocytes,
CC   this provides a good assay system for identifying agonists/antagonists
CC   of C140R. The C140 receptor is a G-protein linked receptor and a member
CC   of the "seven-pass" transmembrane receptor superfamily (peptide chain
CC   of the receptor passes through the cell membrane seven times, producing
CC   seven transmembrane regions within the receptor molecule). The C140
CC   receptor is involved in controlling blood pressure. C140 antagonists
CC   (see WO1942-WO1951) are useful to inhibit signalling from this
CC   receptor, resulting in an increase in blood pressure and are therefore
CC   useful in pharmaceuticals for the treatment of hypotension (low blood
CC   pressure). Conversely agonists (see WO1914-WO1941) of C140 are useful
CC   in pharmaceuticals for the treatment of hypertension (high blood
CC   pressure).
SQ   Sequence   399 AA;

Query Match             21.7%;   Score 423;  DB 1;  Length 399;
Best Local Similarity   31.8%;   Pred. No. 5.3e-39;
Matches  107;   Conservative   55;  Mismatches  141;   Indels   34;   Gaps   10;

Qy     36 ILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLIYY 95
          |:|  ||:| |||:|||:||:|  ||:|||:|||| |:|:||||::|| |:|:|||||
Db     78 VFLPVVYIIVFVIGLPSNGMALWIFLFRTKKHPAVIYMANLALADLLSVIWFPLKISYH 137

Qy     96 AAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPR----L 151
          |:|  |:|  | ||  ||::||:||| || ||:|:|:||: ::|  ||:|| :||
Db    138 LHGNNWVYGEALCKVLIGFFYGNMYCSILFMTCLSVQRYWVIVNPM-----GHPRKKANI 192

Qy    152 AGLLCLAVWLLVVAGCLVP------NLFFVTISNKGTTVLCHDTTRPEEF---DHYVHFSSA 203
          ; |||:|||:|:|:|:||:     ||::||:|||| ||||:|||||||:    |  : ||
Db    193 AVGVSLAIWLLIFLVTIPLYVMKQTIYIPALNITI--CHDVL-PEEVLGVGDMFNYFLSL 248

Qy    204 VMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSRLRSLRTIAVVLTVFAVCFVPFHI 263
          ||  ||| | | |||| |||  |:||| | ||  |::|: :|:  |:|:| ||::||:
Db    249 AIG-VFLPALLTASAYVLMIKTLRSSAMDEHSEKRQRAIRLIITVLAMYFICFAPSNL 307

Qy    264 TRTIYYLARLLEADCRVLNINVNVYKVTRPLASANSCLDPVLYLLTGDKYRRQLRQ--LC 321
          |:|| ||:|| ||:|||  ||| |:|  |  ||  ||:|:|:::|:| :|:|:|     |
Db    308 LLVVHYFLIKTQROSHVY----ALYLVALCLSTLNSCIDPFVYFVYSKDFRDHARNALLC 363

Qy    322 GGGKFQPRTAASSLALVSLPEDSSCRWAATFQDSSCCS 358
          ; |:    |:|::|:    :|
Db    364 RSVRTVNRMQ------ISLSSNKFSRKSGSYSSSSTIS 394
```

```
DT   22-AUG-1995   (first entry)
DE   Murine C140 receptor.
KW   G-protein-coupled receptor; G-protein; C140 receptor.
OS   Mus musculus.
FH   Key               Location/Qualifiers
FT   protein         1..27
FT                   /label= tentative signal sequence
FT   region          78..100
FT                   /label= transmembrane I
FT   region          108..128
FT                   /label= transmembrane II
FT   region          168..191
FT                   /label= transmembrane III
FT   region          168..191
FT                   /label= transmembrane III
FT   region          304..325
FT                   /label= transmembrane VI
FT   region          204..224
FT                   /label= transmembrane IV
FT   region          266..286
FT                   /label= transmembrane V
FT   region          304..325
FT                   /label= transmembrane VI
FT   modified_site   29
FT                   /label= transmembrane VII
FT   modified_site   204..224
FT                   /label= Asn linked glycosylation site
FT   cleavage_site   34..35
FT                   /label= protease receptor cleavage
FT   modified_site   200
FT                   /label= Asn linked glycosylation site
PN   WO9503118-A.
PD   02-FEB-1995.
PF   26-JUL-1994; U08536.
PR   26-JUL-1993; US-097938.
PA   (CORT-) COR THERAPEUTICS.
PI   Scarborough RM, Sundelin J;
WPI; 95-075182/10.
DR   N-PSDB; Q84557.
PN   New DNA encoding recombinant C140 receptor - and novel agonists
PT   and antagonists and specific antibodies with therapeutic and
PT   diagnostic applications.
PS   Disclosure; Fig 1; 57pp; English.
CC   A mouse cosmid genomic library (obtd. from Dr R.A Wetsel, Washington
CC   Univ. School of Medicine, St Louis, Missouri) was screened with two
CC   32P-labeled oligos corresp. to bp 190-249 and 742-801 of the bovine
CC   substance K receptor cDNA. In one of the clones isolated (C140), the
CC   hybridising region was localised to a 3.7 kb PstI fragment. This
CC   fragment was subcloned into pBluescript vector.The hybridising and
CC   adjacent regions were sequenced. The nt sequence and the deduced AA
CC   sequence are given in Q84557 & R69920 respectively.
SQ   Sequence   395 AA;

Query Match             21.6%;   Score 421;  DB 1;  Length 395;
Best Local Similarity   31.8%;   Pred. No. 8.8e-39;
Matches  107;   Conservative   56;  Mismatches  140;   Indels   34;   Gaps   10;

Qy     36 ILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLIYY 95
          |:|  ||:| |||:|||:||:|  ||:|||:|||| |:|:||||::|| |:|:|||||
Db     74 VFLPVVYIIVFVIGLPSNGMALWIFLFRTKKHPAVIYMANLALADLLSVIWFPLKISYH 133

Qy     96 AAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPR----L 151
          |:|  |:|  | ||  ||::||:||| || ||:|:|:||: ::|  ||:|| :||
Db    134 LHGNNWVYGEALCKVLIGFFYGNMYCSILFMTCLSVQRYWVIVNPM-----GHPRKKANI 188

Qy    152 AGLLCLAVWLLVVAGCLVP------NLFFVTISNKGTTVLCHDTTRPEEF---DHYVHFSSA 203
          ; |||:|||:|:|:|:||:     ||::||:|||| ||||:|||||||:    |  : ||
Db    189 AVGVSLAIWLLIFLVTIPLYVMKQTIYIPALNITI--CHDVL-PEEVLGVGDMFNYFLSL 244

Qy    204 VMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSRLRSLRTIAVVLTVFAVCFVPFHI 263
          ||  ||| | | |||| |||  |:||| | ||  |::|: :|:  |:|:| ||::||:
Db    245 AIG-VFLPALLTASAYVLMIKTLRSSAMDEHSEKRQRAIRLIITVLAMYFICFAPSNL 303

Qy    264 TRTIYYLARLLEADCRVLNINVNVYKVTRPLASANSCLDPVLYLLTGDKYRRQLRQ--LC 321
          |:|| ||:|| ||:|||  ||| |:|  |  ||  ||:|:|:::|:| :|:|:|     |
Db    304 LLVVHYFLIKTQROSHVY----ALYLVALCLSTLNSCIDPFVYFVYSKDFRDHARNALLC 359
```

```
RESULT 13
ID   R66920
AC   R66920;
DE   R66920 standard; Protein; 395 AA.
```

```
Qy   322 GGGKPQPRTAASSLALVSLPEDSSCCRWAATPQDSSCS 358
             ||||||||||||||||||||||||||||||||||||||
Db   360 RSVRTVNRMQ------ISLSSNKFSRKSGSYSSSSTS 390

Db   360 RSVRTVNRMQ------ISLSSNKFSRKSGSYSSSSTS 390

RESULT 14
ID   W01952 standard; Protein; 395 AA.
AC   W01952;
DT   01-APR-1997 (first entry)
DE   Murine C140 receptor, including putative signal sequence.
KW   C140 receptor; G-protein linked; coupled; seven pass; agonist;
KW   antagonist; hypertension; blood pressure.
OS   Mus sp.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   peptide         1..27
FT                   /note= "putative signal peptide, differs from
FT                   signal peptide encoded by a cDNA clone of
FT                   this receptor (see W01954); the signal
FT                   sequence given for the cDNA clone is
FT                   believed to be the correct sequence"
FT   protein         28..395
FT                   /note= "mature protein"
FT   modified_site   29
FT                   /note= "potential Asn-linked glycosylation site"
FT   cleavage_site   34..35
FT                   /note= "putative protease receptor cleavage site"
FT   region          78..100
FT                   /note= "transmembrane region I"
FT   region          108..128
FT                   /note= "transmembrane region II"
FT   region          148..169
FT                   /note= "transmembrane region III"
FT   region          188..210
FT                   /note= "transmembrane region IV"
FT   modified_site   220
FT                   /note= "potential Asn-linked glycosylation site"
FT   region          244..264
FT                   /note= "transmembrane region V"
FT   region          286..306
FT                   /note= "transmembrane region VI"
FT   region          324..345
FT                   /note= "transmembrane region VII"
PN   WO9623225-A1.
PD   01-AUG-1996.
PF   25-JAN-1995; U01179.
PR   25-JAN-1995; US-390301.
PA   (CORT-) COR THERAPEUTICS INC.
PS   Scarborough RM, Sundeln J;
PI   Example 1; Fig 1A-B; 60pp; English.
DR   WPI; 96-362833/36.
DR   N-PSDB; T32036.
CC   Vector for expression C140 cell surface receptor in host cell -
CC   useful to identify C140 agonist and antagonists, which are
CC   antihypertensives and elevators of blood pressure, respectively
CC   W01952 represents the murine C140 receptor (C140R), including a putative
CC   signal peptide (see features table). DNA encoding C140R may be
CC   engineered so as to allow the recombinant expression of C140R in a
CC   suitable host cell, i.e. by removing the native expression-control
CC   sequences and replacing them with control sequences operable in the host.
CC   Such a recombinant receptor can be expressed on the surface of oocytes,
CC   this provides a good assay system for identifying agonists/antagonists
CC   of C140R. The C140 receptor is a G-protein linked receptor and a member
CC   of the receptor superfamily (peptide chain
CC   of the "seven-pass" transmembrane receptor passes through the cell
CC   membrane seven times, producing the "seven-pass" receptor molecule). The C140
CC   receptor is involved in controlling blood pressure. C140 antagonists
CC   (see W01942-W01951) are useful to inhibit signalling from this
CC   receptor, resulting in an increase in blood pressure and are therefore
CC   useful in pharmaceuticals for the treatment of hypotension (low blood
CC   pressure). Conversely agonists (see W01914-W01941) of C140 are useful
CC   in pharmaceuticals for the treatment of hypertension (high blood
CC   pressure').
SQ   Sequence    395 AA;
```

```
Qy   36  ILLPYSYAVVFVLGGLGLNAPTLMLFIFRLRPWDATATYMFHLALSDTLLYVLSLPTLIYY 95
             |   | :|     || | |||::|     | |  : :||| |  ||::|   || ||
Db   74  VFLPVVYIIVFVIGLPSNGMALWIFLFRTKKHPAVIYMANLALADLLSVIWFPLKISYH 133

Qy   96  AAHHWHPGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPR----L 151
             |:| :| :|  || ||||||:: | ||:|||| |:| | | || ||| :       |
Db   134 LHGNNWVYGEALCKYLIGFFYGNMYCSILFMTCLSVQRYWVIVNPM------GHPRKKANI 188

Qy   204 VMGLLFGVPCLVTLVCYGLMARRLYQPLPEGSAQSSSRLRSLRRTIAVVLTVFAVCEVPFHI 263
             |:|  : | |  :| :  ||||::|| |     | | ||::| | | | :||| |
Db   189 AVGSSLAIWLILFLVTIPLYVMKQTIYIPALNITT---CHDVL-PEEVLYGDMENYFLSL 244

Qy   152 AGLLCLAVWLLVVAGGLVP----NLFFVTISNKGTTVLCHDTTRPEEF---DHVHFFSA 203

Qy   264 TRTIYYLARLLEADCRVLNIVVNVVYKVTRPLASANSCLDPVLILLTGDKYRYRQLRQ--LC 321
             :|| ||  | |   || :   :   :  | : || | | | | |   | | |  ||
Db   245 AIG-VFLFPALLTRASAYVLMIKTLRSSAMDEHSENKRQRAIRLIITVLAMYFICFRPSNL 303

Qy   304 LLVVHYFLIKTQRQSHVY---ALYILVALCLSTLNSCIDPFVYYFVSKDPRDHARNALLC 359
             |    | :     | | | :    | | |  | :| | |||||||||: | | | |
Db   360 RSVRTVNRMQ------ISLSSNKFSRKSGSYSSSSTS 390
```

```
RESULT 15
ID   R66921 standard; Protein; 398 AA.
AC   R66921;
DT   22-AUG-1995 (first entry)
DE   Human C140 receptor.
KW   G-protein-coupled receptor; G-protein; C140 receptor.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   protein         1..27
FT                   /label= signal peptide
FT   modified_site   31
FT                   /label= Asn linked glycosylation site
FT   modified_site   223
FT                   /label= Asn linked glycosylation site
FT   cleavage_site   37..38
FT                   /label= protease receptor cleavage site
FT   region          81..103
FT                   /label= transmembrane I
FT   region          111..132
FT                   /label= transmembrane II
FT   region          150..174
FT                   /label= transmembrane III
FT   region          191..212
FT                   /label= transmembrane IV
FT   region          245..267
FT                   /label= transmembrane V
FT   region          289..309
FT                   /label= transmembrane VI
FT   region          327..348
FT                   /label= transmembrane VII
PN   WO9503318-A.
PD   02-FEB-1995.
PF   26-JUL-1994; U08536.
PR   26-JUL-1993; US-097938.
PA   (CORT-) COR THERAPEUTICS.
PS   Scarborough RM, Sundeln J;
PI   WPI; 95-075182/10.
DR   N-PSDB; Q84558.
DR   New DNA encoding.recombinant C140 receptor - and novel agonists
```

```
Query Match     Best Local Similarity   21.5%;
Best Local Similarity   31.8%;   Score 418;   DB 1;   Length 395;
Matches 107;   Conservative 55;   Pred. No. 1.9e-38;
             Mismatches 141;   Indels 34;   Gaps 10;
```

```
PT  and antagonists and specific antibodies with therapeutic and
PT  diagnostic applications.
PS  Disclosure; Fig 2; 57pp; English.
CC  The availability of genomic DNA encoding the mouse protease
CC  C140 receptor (see Q84557) permitted the retrieval of the corresp.
CC  human gene. A human genomic library cloned in the vector EMBL3 was
CC  screened using the entire coding region of the murine clone as a
CC  probe. The recovered human gene including the DNA sequence and the
CC  deduced AA sequence are shown in Q84558 & R66921. Subsequent
CC  experiments indicated that the human C140 gene is located in the
CC  same region of the long arm of chromosome number 5 (5q12-5q13)
CC  as has been reported for the human thrombin receptor gene.
SQ  Sequence  398 AA;

Query Match     20.7%;  Score 403;  DB 1;  Length 398;
Best Local Similarity   32.1%;  Pred. No. 9.1e-37;
Matches   95;  Conservative  57;  Mismatches 124;  Indels  20;  Gaps  9;

Qy   36  ILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPLIYY  95
             :|:| |:||:|:|  :|:||:|| ::||:| |  | : | : :|   :  |::|
Db   77  VFLPIVTTIVFVVGLPSNGMALWVFLFRTKKKHPAVIMANLALADLLSVIWFPLKIAYH  136

Qy   96  AAHNHWPPGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLL  155
             |:| : |:  |  |   |:| |||:||:|:|:|||:|| | :| | || ||
Db  137  IHGNNWIYGEALCNVLIGFFYGNMYCSILFMTCLSVQRYWVIVNPMGHSR-KKANIAIGI  195

Qy  156  CLAVWLVVAGCLVP------NLFVTTSNKGTTVLCHDTTRPEEF---DHYVHFSSAVMGL  207
             |:|||:|||:::       :|    | | |   | ||:  ||   |: ||||||:|
Db  196  SLAIWLLILLVTIPLYVVKQTIFIPALNITT---CHDVL-PEQLLVGDMFNYFLSLAIG-  251

Qy  208  LFGVPCLVTLVCYGLMARRLXQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHITRTI  267
             | | :|  || |  | | : |: |   :|  ||:||:|: |||:|| |: :|:  ||: :
Db  251  VFLFPAFLTASAYVLMIRMLRSSAMDENSEKKRKRAIKLIVTVLAMYLICFTPSNLLLVV  310

Qy  268  YYLARLLEADCRVLNIVNVVIKVTRPLASANSCLDPVLYLLTGDKYRRQLRQ--LC  321
             :|  |:| : :|:|: ||  |::  :|| : |::: :|| |:|:: : |    |:
Db  311  HYF--LIKSQGQ--SHVYALYIVALCLSTLNSCIDPFVYYFVSHDFRDHAKNALLC  362
```

Run on: December 29, 1999, 07:20:25 ; Search time 104.56 Seconds
                                              (without alignments)
                                              83.748 Million cell updates/sec

OM nucleic - nucleic search, using sw model

Database :   N_Geneseq_36:*

Searched: 311585 seqs, 125096042 residues

Scoring table: IDENTITY_NUC

Sequence: 1 CAGATCTAGATACTATGTTCTACACTCTTACGTGC 35

Perfect score: 35

Title: US-09-077-173A-3

                                              GenCore version 4.5
                                              Copyright (c) 1993 - 1998 Compugen Ltd.

Word size :   0

Number of hits that pass the threshold :   623170

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	20	57.1	1209	1	T39519	Flea calreticulin
2	20	57.1	1589	1	T39516	Flea calreticulin
3	20	57.1	1209	1	T39517	Flea calreticulin
4	20	57.1	1589	1	T39518	Flea calreticulin
5	19.4	55.4	110000	1	X20248_08	Continuation (9 of
6	19	54.3	1291	1	V59636	Human secreted pro
7	18.6	53.1	5503	1	V18187	Fanconi anaemia of
8	18.6	53.1	9183	1	V60751	Human secreted pro
9	18.6	53.1	1539	1	V60753	HIV-1 strain YBF30
10	18.4	52.6	1990	1	Q33064	HIV-1 strain YBF30
11	18.4	52.6	1990	1	T18995	Encodes Babesia bo
12	18	51.4	846	1	X37420	Babesia merozoite
13	17.8	50.9	150	1	X12163	Human secreted pro
14	17.6	50.3	1192	1	T08635	Human biallelic po
15	17.6	50.3	97	1	T23922	HPV E6-binding pro
16	17.6	50.3	12047	1	V17618	Human gene signatu
17	17.4	49.7	12047	1	X24755	Homo sapiens gland
18	17.4	49.7	10095	1	Q81762	Human glandular ka
19	17.4	49.7	2582	1	X20558	Japanese Black Pin
20	17.4	49.7	1350	1	Q54397	Polynucleotide seq
21	17.2	49.1	1230	1	Q45290	Feline Immunodefic
22	17.2	49.1	1418	1	085885	Sequence of the re
23	17.2	49.1	150	1	085886	gag gene of NCSU1
24	17.2	49.1	892	1	T18979	Open reading frame
25	17.2	49.1	2125	1	T36787	Stanniocalcin alph
26	17.2	49.1	1353	1	T47239	prostate-specific
27	17.2	49.1	23	1	T47249	FIV gag gene. Reco
28	17.2	49.1	1237	1	T74084	FIV gag gene 3'. PC
29	17.2	49.1	1353	1	T49076	Feline immunodefic
30	17.2	49.1	9472	1	V25601	Feline immunodefic
31	17.2	49.1	4727	1	V26971	Consensus vCP255 H
32	17.2	49.1	4616	1	V26971	Feline immunodefic
33	17.2	49.1	2032	1	V26974	Consensus I3L prom
34	17.2	49.1	4150	1	V26976	Consensus vCP329 H
35	17.2	49.1	237326	1	V57903	Hereditary haemoch
36	17.2	49.1	235033	1	V57926	Hereditary haemoch
37	17.2	49.1	9474	1	V58053	FIV F14 clone. Pro
38	17.2	49.1	9474	1	V57355	FIV-NCSU1 clone JS
39	17.2	49.1	9471	1	V57356	FIV-NCSU1 clone JS
40	17	48.6	10723	1	Q12787	Dengue 2 virus gen
41	17	48.6	1349	1	Q57701	K.lactis transaldo
42	17	48.6	1226	1	Q57702	K.lactis transaldo
43	17	48.6	4090	1	Q85027	Entamoeba histolyt
44	17	48.6	133894	1	T13635	AcNPV genomic DNA
45	17	48.6	423	1	T62727	DNA from construct

ALIGNMENTS

RESULT 1
T39519/c
ID   T39519 standard; cDNA; 1209 BP.
AC   T39519;
DT   12-DEC-1996 (first entry)
DE   Flea calreticulin cDNA clone nCtCal1209 (complement).
KW   Calreticulin; flea; haematophagous insect; allergic dermatitis;
KW   vaccine; therapy; nCtcal1209; ss.
OS   Ctenocephalides felis.
PN   WO9628469-Al.
PD   19-SEP-1996.
PF   08-MAR-1996; U03133.
PR   09-MAR-1995; US-401509.
PA   (HESK-) HESKA CORP.
PI   Rushlow KE, Stiegler GL;
PT   Haematophagous insect calreticulin protein - used to reduce insect
PT   infestation and desensitise patients to allergic dermatitis
PS   Claim 16; Page 72; 86pp; English.
CC   The antisense strand (T39519) of nucleic acid nCtCal1209 (T39517)
CC   hybridises under stringent conditions to a flea calreticulin gene.
CC   nCtCal1209 codes for a flea calreticulin protein (W04171) that can
CC   be used to alter the blood feeding behaviour of haematophagous
CC   insects or to desensitise an animal to allergic dermatitis. The
CC   antisense strand, when administered to an animal, reduces
CC   calreticulin activity in insects feeding on the animal, thereby
CC   reducing the insect burden on the animal and in the environment
CC   of the animal.
SQ   Sequence 1209 BP;   292 A;   280 C;   212 G;   425 T;

Qy   1 CAGATCTAGATACTATGTTCTACACTCT 28
               |||||||||||||||||||||||||||
Db   1007 CAGGTCAAGATCCTAGGTTCTACGCTCT 980

Query Match   57.1%; Score 20; DB 1; Length 1209;
Best Local Similarity 82.1%;   Pred. No. 5.7;
Matches   23; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

RESULT 2
T39516
ID   T39516 standard; cDNA; 1589 BP.
AC   T39516;
DT   12-DEC-1996 (first entry)
DE   Flea calreticulin cDNA clone nCtCal1589.
KW   Calreticulin; flea; haematophagous insect; allergic dermatitis;
KW   vaccine; therapy; nCtcal1589; ss.
OS   Ctenocephalides felis.
FH   Key   Location/Qualifiers
FT   cds   151..1362
FT   /*tag= a
PN   WO9628469-Al.
PD   19-SEP-1996.
PF   08-MAR-1996; U03133.
PR   09-MAR-1995; US-401509.
PA   (HESK-) HESKA CORP.
PI   Rushlow KE, Stiegler GL;
P-PSDB: W04171.
DR   96-44261/44.
PT   Haematophagous insect calreticulin protein - used to reduce insect
PT   infestation and desensitise patients to allergic dermatitis

```
PS   Claim 7; Page 66-68; 86pp; English.
CC   A cDNA clone (T39516), designated nCtcal1589, codes for a flea
CC   calreticulin protein (W04171) that can be used to alter the blood
CC   feeding behaviour of haematophagous insects or to desensitise an
CC   animal to allergic dermatitis. Its sequence was deduced from
CC   clones nCtcal1665, nCtcal1750 and nCtcal1218 isolated from a
CC   Ctenocephalides felis salivary gland cDNA library by PCR
CC   amplification using flea calreticulin gene-specific primers (see
CC   also T39529-32).
CC   nCtcal1589 and other calreticulin nucleic acids
CC   (see also T39515 and T39517-20) can be administered to an animal
CC   against haematophagous insect infestation.
SQ   Sequence   1589 BP;   552 A;   266 C;   335 G;   436 T;

Qy    1 CAGATCTAGATACTATGTTCTCACACTCT 28
        ||| || |||| ||| ||||||||| ||||
ob  353 CAGGTCAAGATGCTAGGTTCTACGCTCT 380
```

```
Query Match   57.1%;   Score 20;   DB 1;   Length 1589;
Best Local Similarity   82.1%;   Pred. No. 6;
Matches  23;   Conservative   0;   Mismatches   5;   Indels   0;   Gaps   0;
```

```
RESULT 3
T39517
ID   T39517 standard; cDNA; 1209 BP.
AC   T39517;
DT   12-DEC-1996 (first entry)
DE   Flea calreticulin cDNA clone nCtcal1209.
KW   calreticulin; flea; haematophagous insect; allergic dermatitis;
KW   vaccine; therapy; nCtcal1209; ss.
OS   Ctenocephalides felis.
PN   WO9628469-A1.
PD   19-SEP-1996.
PF   08-MAR-1995; 09-MAR-1995; U03133.
PR   09-MAR-1995; US-401509.
PA   (HESK-) HESKA CORP.
PI   Rushlow KE, Stiegler GL;
DR   WPI; 96-442861/44.
DR   P-PSDB; W04171.
PT   Haematophagous insect calreticulin protein - used to reduce insect
PT   infestation and desensitise patients to allergic dermatitis
PS   Claim 7; Page 71; 86pp; English.
CC   Nucleic acid nCtcal1209 (T39517) comprises the coding region (minus
CC   the stop codon) encoding flea calreticulin PCtcal1403 (W04171), a
CC   protein that can be used to alter the blood feeding behaviour of
CC   haematophagous insects or to desensitise an animal to allergic
CC   dermatitis.
CC   The coding sequence was deduced from a cDNA clone
CC   (T39516) isolated from a cDNA library prepd. from the salivary
CC   glands of Ctenocephalides felis. Calreticulin nucleic acids (see
CC   also T39515 and T39518-20) can be used to produce recombinant
CC   calreticulin, or can be administered to an animal to protect it
CC   against haematophagous insect infestation.
SQ   Sequence   1209 BP;   425 A;   212 C;   280 G;   292 T;
```

```
Query Match   57.1%;   Score 20;   DB 1;   Length 1589;
Best Local Similarity   82.1%;   Pred. No. 6;
Matches  23;   Conservative   0;   Mismatches   5;   Indels   0;   Gaps   0;

Qy    1 CAGATCTAGATACTATGTTCTCACACTCT 28
        ||| || |||| ||| ||||||||| ||||
Db  203 CAGGTCAAGATGCTAGGTTCTACGCTCT 230
```

```
RESULT 4
T39518/c
ID   T39518 standard; cDNA; 1589 BP.
AC   T39518;
DT   12-DEC-1996 (first entry)
DE   Flea calreticulin cDNA clone nCtcal1589 (complement).
KW   calreticulin; flea; haematophagous insect; allergic dermatitis;
KW   vaccine; therapy; nCtcal1589; ss.
OS   Ctenocephalides felis.
PN   WO9628469-A1.
PD   19-SEP-1996.
PF   08-MAR-1995; 09-MAR-1995; U03133.
PR   09-MAR-1995; US-401509.
PA   (HESK-) HESKA CORP.
PI   Rushlow KE, Stiegler GL;
     WPI; 96-442861/44.
     Claim 16; Page 70; 86pp; English.
CC   Haematophagous insect calreticulin protein - used to reduce insect
CC   infestation and desensitise patients to allergic dermatitis
CC   The antisense strand (T39518) of nucleic acid nCtcal1589 (T39516)
CC   hybridises under stringent conditions to a flea calreticulin gene.
CC   nCtcal1589 codes for a flea calreticulin protein (W04171) that can
CC   be used to alter the blood feeding behaviour of haematophagous
CC   insects or to desensitise an animal to allergic dermatitis. The
CC   antisense strand, when administered to an animal, reduces
CC   calreticulin activity in insects feeding on the animal, thereby
CC   reducing the insect burden on the animal and in the environment
CC   of the animal.
SQ   Sequence   1589 BP;   436 A;   335 C;   266 G;   552 T;
```

```
Query Match   57.1%;   Score 20;   DB 1;   Length 1589;
Best Local Similarity   82.1%;   Pred. No. 6;
Matches  23;   Conservative   0;   Mismatches   5;   Indels   0;   Gaps   0;

Qy    1 CAGATCTAGATACTATGTTCTCACACTCT 28
        ||| || |||| ||| ||||||||| ||||
Db 1237 CAGGTCAAGATGCTAGGTTCTACGCTCT 1210
```

```
RESULT 5
X20248_08/c
Continuation (9 of 10) of X20248 from base 800001
WP Sequence split into 10 fragments   LOCUS X20248 Accession X20248
                                      (Borrelia burgdorferi polynucleotid
```

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

```
Query Match   55.4%;   Score 19.4;   DB 1;   Length 110000;
Best Local Similarity   79.3%;   Pred. No. 21;
Matches  23;   Conservative   0;   Mismatches   6;   Indels   0;   Gaps   0;

Qy      1 CAGATCTAGATACTATGTTCTACACTCTT 29
          ||||| ||||| || ||| |||||||| ||
Db  22738 CAGATAAAGATAATATTTTGTACACTATT 22710
```

```
RESULT 6
V59636
ID   V59636 standard; DNA; 1291 BP.
AC   V59636;
DT   19-JAN-1999 (first entry)
DE   Human secreted protein gene 126 clone HELGH31.
KW   Human; secreted protein; fusion protein; gene therapy; protein
KW   therapy; diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
```

```
OS   Homo sapiens.
PN   WO9839448-A2.
PD   11-SEP-1998;
PF   06-MAR-1997;
PF   02-OCT-1997; US-061060.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.
PR   11-APR-1997; US-043669.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   11-APR-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047581.
PR   23-MAY-1997; US-047588.
PR   23-MAY-1997; US-047587.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047589.
PR   23-MAY-1997; US-047590.
PR   23-MAY-1997; US-047592.
PR   23-MAY-1997; US-047593.
PR   23-MAY-1997; US-047594.
PR   23-MAY-1997; US-047595.
PR   23-MAY-1997; US-047596.
PR   23-MAY-1997; US-047597.
PR   23-MAY-1997; US-047598.
PR   23-MAY-1997; US-047599.
PR   23-MAY-1997; US-047600.
PR   23-MAY-1997; US-047601.
PR   23-MAY-1997; US-047612.
PR   23-MAY-1997; US-047613.
PR   23-MAY-1997; US-047614.
PR   23-MAY-1997; US-047615.
PR   23-MAY-1997; US-047617.
PR   23-MAY-1997; US-047618.
PR   23-MAY-1997; US-047632.
PR   23-MAY-1997; US-047633.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048974.
PR   13-JUN-1997; US-049610.
PR   08-JUL-1997; US-051926.
PR   16-JUL-1997; US-052874.
PR   18-AUG-1997; US-055724.
PR   22-AUG-1997; US-055630.
PR   22-AUG-1997; US-055631.
PR   22-AUG-1997; US-056632.
PR   22-AUG-1997; US-056636.
PR   22-AUG-1997; US-056637.
PR   22-AUG-1997; US-056662.
PR   22-AUG-1997; US-056664.
PR   22-AUG-1997; US-056845.
PR   22-AUG-1997; US-056862.
PR   22-AUG-1997; US-056864.
PR   22-AUG-1997; US-056872.
PR   22-AUG-1997; US-056874.
PR   22-AUG-1997; US-056875.
PR   22-AUG-1997; US-056876.
PR   22-AUG-1997; US-056877.
PR   22-AUG-1997; US-056878.
PR   22-AUG-1997; US-056879.
PR   22-AUG-1997; US-056880.
PR   22-AUG-1997; US-056881.
PR   22-AUG-1997; US-056882.
PR   22-AUG-1997; US-056884.
PR   22-AUG-1997; US-056886.
PR   22-AUG-1997; US-056887.
PR   22-AUG-1997; US-056888.
PR   22-AUG-1997; US-056889.
PR   22-AUG-1997; US-056892.
PR   22-AUG-1997; US-056893.
PR   22-AUG-1997; US-056894.
PR   22-AUG-1997; US-056903.
PR   22-AUG-1997; US-056908.
PR   22-AUG-1997; US-056909.
PR   22-AUG-1997; US-056910.
PR   22-AUG-1997; US-056911.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057669.
PR   05-SEP-1997; US-057761.
PR   12-SEP-1997; US-058785.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI   Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI   Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI   Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
PT   New isolated human genes, and the secreted polypeptide(s) they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders.
DR   P-PSDB: W74854.
DR   P-PSDB: W74854.
WPI: 98-506364/43.
CC   This sequence represents a nucleic acid molecule designated Gene 126 from
CC   the human cDNA clone HELGH31 (deposited as clone ATCC 97902 and ATCC
CC   209048) which encodes a secreted human protein. The gene can be used to
CC   generate fusion proteins by linking to the gene to a human immunoglobulin
CC   Fc portion (e.g. V59502) for increasing the stability of the fused
CC   protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polypeptides, based on
CC   which tissues they are most highly expressed in (see V59511 for described
CC   uses).
PS   Claim 1; Page 360-361; 721pp; English.
SQ   Sequence 1291 BP; 391 A; 228 C; 203 G; 469 T;

Query Match             54.3%;  Score 19;  DB 1;  Length 1291;
Best Local Similarity   71.4%;
Matches   25;  Conservative   0;  Mismatches  10;  Indels   0;  Gaps   0;

Qy          1 CAGATCTAGATACTATGTTCTACACTCTTACGTGC 35
              |||||  | |  || ||||||||| |||| || ||
Db       1015 CAGATTCAAAATGTGTGTTCTACACTGTTACAGGC 1049

RESULT   7
ID   V18187
AC   V18187;
DT   28-AUG-1998 (first entry)
ID   V18187 standard; cDNA to mRNA; 5503 BP.
```

```
DE   Fanconi anaemia of complementation group A gene.
KW   Fanconi anaemia of complementation group A; FA-A; genetic defect;
KW   prenatal FA-A; FA-A carrier detection; disease diagnosis; ss.
OS   Homo sapiens.
FH   Key          Location/Qualifiers
FT   CDS          32..4399
FT                /product= a
PN   W09814462-A1.
PD   09-APR-1998.
PF   03-OCT-1997; U18010.
PR   04-OCT-1996; US-726012.
PA   (FANC-) FANCONI ANEMIA RES FUND INC.
PI   Joenje H, Lo Ten Foe JR.
     WPI; 98-240012/21.
PD   12-JUN-1998.
DR   FR2756843-A1.
DR   P-PSDB; W48663.
PT   DNA for Fanconi Anaemia complementation group A for diagnosis and
     developing products for diagnosis and screening of disease and gene
PT   therapy.
PS   Claim 4; Fig 2; 63pp; English.
CC   This sequence encodes the Fanconi anaemia of complementation group A
CC   (FA-A) protein of the invention. The DNA's may be used to complement a
CC   genetic defect in a cell (especially the FA-A gene). The products can be
CC   used for screening (especially prenatal FA-A), detection of FA-A carriers
CC   and FA-A disease diagnosis.
SQ   Sequence 5503 BP; 1208 A;     1527 C;     1492 G;     1276 T;

Qy     2 AGATCTAGATACTATGTTCTACACTCTTACGTG 34
          |||||| ||| | ||||| | |||||| ||||||
Db  5354 AGATCTTTAAACTGCTTTATACACTGTACACGTG 5386

Query Match            53.1%;  Score 18.6;  DB 1;  Length 5503;
Best Local Similarity  72.7%;
Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

RESULT 8
ID   V60751 standard; DNA; 9183 BP.
AC   V60751;
DT   08-DEC-1998  (first entry)
DE   HIV-1 strain YBF30 complete genomic sequence.
KW   HIV-1 strain YBF30; antibody; oligonucleotide; diagnosis; immunisation;
OS   Human immunodeficiency virus type 1.
PN   FR2756843-A1.
PD   12-JUN-1998.
PF   09-DEC-1996; 015087.
PR   09-DEC-1996; FR-015087.
PA   (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
PI   Barre-Sinoussif F, Loussert-Ajaka I, Mauclere P, Saragosti S,
PI   Simon F;
     WPI; 98-336114/30.
PT   Non-M, non-O HIV-1 strain YBF30 - useful for diagnosis and
PT   immunisation
PS   Claim 3; Page 19-24; 85pp; French.
CC   This sequence represents the complete genome sequence of the non-M
CC   (major), non-O (Outlier) HIV-1 strain YBF30 (CNCM I-1753), isolated from
CC   the Cameroon. The HIV strain, peptides, antibodies and oligonucleotides
CC   derived from it (see V60752-V60798 and W68473-W68482) are used for
CC   diagnosis of or immunisation against non-M, non-O HIV-1 infections.
CC   The oligonucleotides, peptides and antibodies can also be used for
CC   typing HIV strains.
SQ   Sequence 9183 BP; 3269 A;     1640 C;     2227 G;     2047 T;

Query Match            53.1%;  Score 18.6;  DB 1;  Length 9183;
Best Local Similarity  84.0%;  Pred. No. 32;
Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy     1 CAGATCTAGATACTATGTTCTACAC 25
          ||||||||| |||||||||||| | ||
Db   899 CAGATCTAAAATACTATGTTAAATAC 923

RESULT 9
ID   V60753 standard; DNA; 1539 BP.
AC   V60753;
DT   08-DEC-1998  (first entry)
DE   HIV-1 strain YBF30 gag gene.
KW   HIV-1 strain YBF30; antibody; oligonucleotide; diagnosis; immunisation;
OS   Human immunodeficiency virus type 1.
FH   Key          Location/Qualifiers
FT   CDS          1..1539
FT                /tag= a
FT                /product= "gag protein"
PN   FR2756843-A1.
PD   12-JUN-1998.
PF   09-DEC-1996; 015087.
PR   09-DEC-1996; FR-015087.
PA   (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
PI   Barre-Sinoussif F, Loussert-Ajaka I, Mauclere P, Saragosti S,
PI   Simon F;
     WPI; 98-336114/30.
PT   Non-M, non-O HIV-1 strain YBF30 - useful for diagnosis and
PT   immunisation
PS   Claim 3; Page 25-27; 85pp; French.
CC   This sequence represents the gag gene from the non-M (major), non-O
CC   (Outlier) HIV-1 strain YBF30 (CNCM I-1753), isolated from the Cameroon.
CC   The HIV strain (see V60751 for complete genome), peptides, antibodies and
CC   oligonucleotides derived from it (see V60752-V60798 and W68473-W68482)
CC   are used for diagnosis of or immunisation against non-M, non-O HIV-1
CC   infections. The oligonucleotides, peptides and antibodies can also be
CC   used for typing HIV strains.
SQ   Sequence 1539 BP; 531 A;     314 C;     292 G;     402 T;

Query Match            53.1%;  Score 18.6;  DB 1;  Length 1539;
Best Local Similarity  84.0%;  Pred. No. 24;
Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy     1 CAGATCTAGATACTATGTTCTACAC 25
          ||||||||| ||||||||||| | ||
Db   551 CAGATCTAAAATACTATGTTAAATAC 575

RESULT 10
ID   Q33064 standard; DNA; 1990 BP.
AC   Q33064;
DT   06-MAY-1993  (first entry)
DE   Encodes Babesia bovis 60kD immunoreactive merozoite surface epitope.
KW   babesiosis; cows; cattle; bos taurus; babesia bovis; babesia bigemina;
KW   merozoite; schizont; ss.
OS   Babesia bovis.
FH   Key          Location/Qualifiers
FT   cds          1..1990
FT                /tag= a
PN   US5171685-A.
PD   15-DEC-1992.
PF   04-APR-1990; 504461.
PR   04-APR-1990; US-504461.
PA   (UYFL ) UNIV FLORIDA.
PA   (USDA ) US SEC OF AGRIC.
PI   Davis WC, Goff WL, Hines SA, Jasmer DP, Mcelwain TF;
PI   Mcguire TC, Palmergh, Perryman LE, Reduker DW;
     WPI; 93-008592/01.
DR   P-PSDB; R30613;
```

PT   DNA encoding Babesia bovis protein 7 - is used as probes and for
PT   prodn. of polypeptide(s) for use in vaccines and for prodn. of
PT   antibodies
PS   Example 19: Fig 3; 20pp; English.
CC   This sequence encodes an immunoreactive epitope located on the
CC   surface of babesia bovis merozoites. A. B. bovis cDNA expression
CC   library was constructed using poly A(+). RNA isolated from B. bovis
CC   Infected blood cultures. Plaques were immunoscreened using rabbit
CC   developmental abnormality; and positive plaques tested for reactivity with
CC   anti-Bv60 sera, and positive plaques tested for reactivity with
CC   monoclonal antibodies that recognised a Bv42 surface exposed epitope
CC   as well as an isotype control monoclonal antibody and normal rabbit
CC   serum. Lambda rBv60 phagemid DNA was isolated from bacteria, and
CC   then restriction digested.
SQ   Sequence 1990 BP; 628 A; 437 C; 398 G; 527 T;

Qy   5   TCTAGAGACTAAGTTCTACACACTCTTACG 32
      |||||||||||||||||||||| || ||| |||
Db   79   TTTAGAGACTAAGTTCAATAAATTACG 106

Query Match   52.6%;   Score 18.4;   DB 1;   Length 1990;
Best Local Similarity   78.6%;   Pred. No. 30;
Matches   22;   Conservative   0;   Mismatches   6;   Indels   0;   Gaps   0;

RESULT 11
ID   T18995
ID   T18995; standard; cDNA; 1990 BP.
AC   T18995;
DT   15-OCT-1996 (first entry)
DE   Babesia merozoite surface protein cDNA clone Bv60.
KW   Babesiosis; merozoite surface protein; vaccine; probe; diagnosis; ss.
OS   Babesia bovis.
FH   Key   Location/Qualifiers
FT   cds.   122..1819
FT   /*tag= a
PN   US5518916-A.
PD   21-MAY-1996.
PF   04-APR-1989; 333155.
PF   04-APR-1989; US-333155.
PR   04-APR-1990; US-504461.
PR   14-DEC-1992; US-989616.
PR   21-NOV-1994; US-342480.
PA   (USDA) US SEC OF AGRIC.
PI   Goff WL, Jasmer DP, McElwain TF, McGuire TC, Reduker DW;
PI   Stiller D;
WPI   96-259067/26.
P-PSDB; R97981.
PT   New fragment of Babesia bovis genomic DNA - useful as a probe for
PT   detecting Babesia infection
PS   Example 18: Fig 3; 19pp; English.
CC   A cDNA clone (T18995) Codes for Bv60, a 60 kDa immunoreactive
CC   protein (R97981) located on the surface of Babesia bovis merozoites.
CC   It was isolated from a blood-stage B. bovis cDNA library in lambda
CC   ZAPII by subcloning into Bluescript SK(-) and immunoscreening using
CC   monospecific anti-Bv60 antisera. Bv60, Bo44 and Bv42 (see also
CC   T18993 and T18994) DNA sequences can be used to make recombinant
CC   Proteins useful as vaccines for the prophylaxis of bovine babesiosis.
CC   They can also be used as diagnostic probes.
SQ   Sequence 1990 BP; 628 A; 437 C; 398 G; 527 T;

Qy   5   TCTAGAGACTATGTTCTACACACTCTTACG 32
      ||||||||||| ||| |||| | | |||||| |
Db   79   TTTAGAGACTAAGTTCAATAAATTACG 106

Query Match   52.6%;   Score 18.4;   DB 1;   Length 1990;
Best Local Similarity   78.6%;   Pred. No. 30;
Matches   22;   Conservative   0;   Mismatches   6;   Indels   0;   Gaps   0;

RESULT 12
X37420

Qy   5   TCTAGAGACTATGTTCTACACACTCTTACG 32
      |||||||||||||||||||| | | || ||||
Db   79   TTTAGAGACTAAGTTCAATAAATTACG 106

ID   X37420 standard; cDNA; 846 BP.
AC   X37420;
DT   06-JUL-1999 (first entry)
DE   Human secreted protein cDNA fragment containing gene 52.
DE   Human; secreted protein; prevention; treatment; protein therapy;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; foetal deficiency; blood disorder; lymphoma;
KW   leukemia; immune system disorder; autoimmune disease; hepatic disease;
KW   renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;
KW   Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;
KW   arthritis; psoriasis; digestive; endocrine; infection; ss.
OS   Homo sapiens.
PN   WO9909155-A1.
PD   25-FEB-1999.
PF   18-AUG-1998; U17044.
PR   16-JUN-1998; US-092956.
PR   15-JUL-1998; US-092956.
PR   19-AUG-1997; US-056368.
PR   19-AUG-1997; US-053369.
PR   19-AUG-1997; US-056535.
PR   19-AUG-1997; US-056555.
PR   19-AUG-1997; US-056556.
PR   19-AUG-1997; US-056628.
PR   19-AUG-1997; US-056629.
PR   19-AUG-1997; US-056726.
PR   19-AUG-1997; US-056728.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Brewer LA, Duan R, Ebner R, Endress GA, Feng P,
PI   Florence C, Florence KA, Komatsoulis GA, Lafleur DW,
PI   Moore PA, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR,
PI   Young PE;
WPI   99-190160/16.
P-PSDB; Y07795.
DR   P-PSDB: Y07795.
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Claim 1a; Page 219; 280pp; English.
CC   This invention describes novel isolated human secreted proteins and
CC   their encoding nucleic acid sequences. The products of the invention
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also pathological conditions can be
CC   diagnosed by determining the presence or amount of expression of
CC   the new polypeptides in a sample or by determining the presence or
CC   absence of mutations in the new polynucleotides. Specific uses are
CC   described for each of the 70 polynucleotides, based on which tissues they
CC   are most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, tumours, neurodegenerative
CC   disorders, developmental abnormalities and foetal deficiencies, blood
CC   disorders, leukemias, diseases of the immune system, autoimmune diseases,
CC   hepatic and renal disease, lymphomas, inflammation, allergies, asthma,
CC   sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,
CC   osteoporosis, arthritis, psoriasis, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in X07744-Y07850 and the encoding nucleic acids are
CC   represented in X37369-X37441.
SQ   Sequence 846 BP; 246 A; 147 C; 189 G; 264 T;

Qy   1   CAGATCTAGATACTATGTTCTACACTCTTACGTG 34
      |||| || | |||| | |||| ||||||| |||||||
Db   80   CGGAACTATAAAAAATGTATTCACTTTACGTG 113

Query Match   51.4%;   Score 18;   DB 1;   Length 846;
Best Local Similarity   70.6%;   Pred. No. 39;
Matches   24;   Conservative   0;   Mismatches   10;   Indels   0;   Gaps   0;

RESULT 13
X12163/c

ID   X12163 standard; DNA; 150 BP.
AC   X12163;
DT   30-MAR-1999 (first entry)
DE   Human biallelic polymorphic DNA fragment EST365190a.

```
KW  polymorphism; biallelic; human; forensic; paternity testing; disease;
KW  detection; phenotypic typing; characteristic; infection; hereditary;
KW  autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW  treatment; marker; ss.
OS  Homo sapiens.
PN  WO9820165-A2.
PD  14-MAY-1998.
PR  05-NOV-1997; US-030455.
PR  06-NOV-1996; US-030455.
PI  Hudson T, Lander ES, Wang D;
PA  (WHED ) WHITEHEAD INST BIOMEDICAL RES.
DR  WPI; 98-286974/25.
PT  New isolated nucleic acid segments from the human genome - used for
PT  determining polymorphic forms for use in e.g. forensics, paternity
PT  testing or phenotypic typing for disease
PS  Claim 1; Page 225; 310pp; English.
CC  X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC  markers which have been isolated using the primers represented in
CC  X09121-X10268. The base occupying the polymorphic site is indicated by
CC  the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC  methods for determining polymorphic forms in an individual for use in
CC  e.g. forensics, paternity testing or for phenotypic typing for diseases
CC  such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC  muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC  hypercholesterolemia, polycystic kidney disease, hereditary
CC  spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC  haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC  syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC  autoimmune diseases, inflammation, cancer, diseases of the nervous
CC  system, infection by pathogenic microorganisms, and characteristics such
CC  as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC  endurance, fertility, and susceptibility to particular
CC  drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC  segments can also be used to produce medicaments for the treatment or
CC  prophylaxis of such diseases.
SQ  Sequence  150 BP;  34 A;  53 C;  30 G;  32 T;

Query Match            50.9%;       Score 17.8;  DB 1;  Length 150;
Best Local Similarity  75.9%;       Pred. No. 36;
Matches  22;  Conservative  0;  Mismatches  7;  Indels  0;  Gaps  0;

Qy     7  TAGAATACTAATGTTCTACACTCTTACGTGC  35
          ||| || |||| | ||||| | |||| ||||
Db    89  TAGGTACAAAGTTGTCCAGTCTTAGGTGC  61
```

```
RESULT  14
ID   T08635  standard; cDNA; 1192 BP.
AC   T08635;
DT   15-JUN-1995; U07780.
PR   08-JUL-1994; US-273059.
DT   08-MAY-1996  (first entry)
DE   HPV E6-binding protein cDNA clone SD-7.
KW   E6 binding protein; E6-BP; human papilloma virus; HPV; antagonist;
KW   diagnosis; ss.
OS   Homo sapiens.
FH   Key              Location/Qualifiers
FT   cds              173..832
FT                    /*tag= a
PN   WO9602000-A1.
PD   25-JAN-1996.
PR   15-JUN-1995; U07780.
PR   08-JUL-1994; US-273059.
PA   (NEWE-) NEW ENGLAND MED CENT HOSPITALS.
PI   Androphy E, Chen JJ;
DR   WPI; 96-097705/10.
DR   P-PSDB; R77660.
PT   Papilloma virus E6-binding proteins - used to develop prods. for the
PT   study, diagnosis and treatment of papilloma virus infection
PS   Claim 11; Page 53; 71pp; English.
CC   A cDNA clone (T08635), designated SD-7, codes for the N-terminal
CC   portion (R77660) of E6-BPSD-7, a protein that specifically binds to
```

```
CC   human papilloma virus (HPV) protein E6.  SD-7 and other E6-BP-encoding
CC   clones (see T08635-41) were identified using a two-hybrid assay in
CC   which yeast expressing the HPV-16 E6 gene fused to the BPV E2 DNA-
CC   binding domain was transformed with a vector carrying HeLa cDNA and a
CC   VP16 transcription activation domain. These novel cDNA clones
CC   (deposited as a library of pRS306 plasmids as ATCC 75827) are used to
CC   produce recombinant E6-BPs in host cells, as primers or probes e.g. to
CC   detect genetic lesions, to raise transgenic animals, and in gene
CC   therapy.
SQ   Sequence  1192 BP;  382 A;  191 C;  260 G;  359 T;

Query Match            50.3%;       Score 17.6;  DB 1;  Length 1192;
Best Local Similarity  71.9%;       Pred. No. 62;
Matches  23;  Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;

Qy     2  AGAATCTAGAATACTAATGTTCTACACTCTTACGT  33
          | ||| ||| || ||||||| ||||||||||| ||
Db   901  AAAATATTGATGTTGTATTTTACACTCTTAAGT  932
```

```
RESULT  15
ID   T23922  standard; cDNA to mRNA; 97 BP.
AC   T23922;
DT   23-AUG-1996  (first entry)
DE   Human gene signature HUMGS05862.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PR   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206991/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1487; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence  97 BP;  26 A;  17 C;  34 G;  17 T;

Query Match            50.3%;       Score 17.6;  DB 1;  Length 97;
Best Local Similarity  71.9%;       Pred. No. 41;
Matches  23;  Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;

Qy     3  GATCTAGATACTAATGTTCTACACTCTTACGTG  34
          |||| | |||| ||||||| |||||| || |||
Db     1  GATCTTPAAACTGCTTTATACACTGTCACGTG  32
```

Search completed: December 29, 1999, 07:20:31

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 07:20:31 ; Search time 104.56 seconds
                                              (without alignments)
                                              83.748 Million cell updates/sec

                 GenCore version 4.5
                 Copyright (c) 1993 - 1998 Compugen Ltd.

Title: US-09-077-173A-4
Perfect score: 35
Sequence: 1 TCTTAAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database :   N_Geneseq_36:*

Word size :   0

Number of hits that pass the threshold :   623170

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19.6	56.0	237326	1	V57903	Hereditary haemoch
2	19.6	56.0	235033	1	V57926	Hereditary haemoch
3	18	51.4	1927	1	Q47841	Flavonoid-3',5'-hy
4	18	51.4	2600	1	Q90096	Mouse kappa-3 opio
5	17.8	50.9	574	1	T19014	Human gene signatu
6	17.6	50.3	1537	1	T26895	Bacteriophage resi
7	17.6	50.3	3018	1	V84620	Human secreted pro
8	17.6	50.3	3735	1	V84429	Human secreted pro
9	17.6	50.3	4005	1	V74942	Staphylococcus aur
10	17.4	49.7	2078	1	T47698	Banana polyphenol
11	17.2	49.1	5433	1	Q40945	Human DNA polymera
12	17.2	49.1	138894	1	T13635	AcNPV genomic DNA
13	17.2	49.1	984	1	T13636	AcNPV ORF 13, resi
14	17.2	49.1	1679	1	V03349	Homo sapiens inter
15	17.2	49.1	4854	1	V74417	Staphylococcus aur
16	17	48.6	17410	1	Q53142	Sequence encoding
17	17	48.6	17350	1	Q72708	hOP1 human osteoge
18	17	48.6	9487	1	Q79140	Hepatitis C virus
19	17	48.6	17415	1	T11639	Human osteogenic P
20	17	48.6	17410	1	T18381	hOP-1 genomic DNA.
21	17	48.6	6595	1	T94659	Arabidopsis flavon
22	17	48.6	17410	1	V15205	Human osteogenic P
23	17	48.6	9409	1	V5294	Streptococcus pneu
24	17	48.6	17415	1	V80733	Human osteogenic p
25	17	48.6	17410	1	X00230	Human osteogenic p
26	17	48.6	616	1	X21012	Polynucleotide seq
27	16.6	47.4	134525	1	Q04525	Total base sequenc
28	16.6	47.4	134525	1	Q04525	Total base sequenc
29	16.6	47.4	1900	1	N70813	Aspergillus nidula
30	16.6	47.4	402	1	Q60506	Human brain Expres
31	16.6	47.4	406	1	Q68478	T cell receptor Va
32	16.6	47.4	1938	1	T71312	DNA encoding threo
33	16.6	47.4	5253	1	T64877	Neuroblastoma indi
34	16.6	47.4	5595	1	T64850	Microtubule-associ
35	16.6	47.4	1938	1	V48127	Staphylococcus aur
36	16.6	47.4	3111	1	V54586	Human secretory pr
37	16.6	47.4	13518	1	X20563	Polynucleotide seq
38	16.6	47.4	1380	1	X24100	Plasmid pBBP21 DNA
39	16.6	47.4	1219	1	X24099	Plasmid pBBP20 DNA
40	16.6	47.4	793	1	X24101	Plasmid pBBP22 DNA
41	16.4	46.9	3744	1	Q10157	Sequence contg. ab
42	16.4	46.9	5394	1	T00872	Murine mC26 struct
43	16.4	46.9	3269	1	T31291	Rat poly-immunoglo
44	16.4	46.9	110000	1	T58810_2	Continuation (3 of
45	16.4	46.9	1366	1.	T67284	Human urogenital s

ALIGNMENTS

RESULT   1
V57903
ID   V57903 standard; DNA; 237326 BP.
AC   V57903;
DT   21-DEC-1998 (first entry)
DE   Hereditary haemochromatosis subregion from an HH affected individual.
KW   Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW   diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW   BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW   type 1 sodium transport gene; ss.
OS   Homo sapiens.
PN   WO9814466-A1.
PD   09-APR-1998.
PF   30-SEP-1997; U17658.
PR   07-MAY-1997; US-852495.
PR   01-OCT-1996; US-724394.
PA   (PROG-) PROGENTIOR INC.
PI   Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI   Tsuchihashi Z, Wolff RK;
DR   WPI; 98-240014/21;
PT   Hereditary haemochromatosis gene products - used to develop products
PT   for the diagnosis and treatment of hereditary disorders in iron
PT   metabolism
PS   Claim 1; Fig 9; 209pp; English.
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. The present sequence
CC   represents a hereditary haemochromatosis subregion from an hereditary
CC   haemochromatosis (HH) affected individual. Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   presence of the HFE gene mutation in the genome of the individual. The
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes BTF genes, which are homologues of the milk
CC   protein butyrophilin (BT), and can be used in the production of agonists
CC   and antagonists of BT function. Also described are: (1) a RoRet gene
CC   which can be used to develop products for the study, diagnosis and
CC   treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC   which are homologues of a type 1 sodium transport gene, and can
CC   similarly be used for hypophosphatemia.
SQ   Sequence 237326 BP;

Qy   2 CTTAAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35
            |||||||||||||||||| ||||||||||||||||
Db   83534 CTAAAGCTTGGTGTCATTAAGGAGATGTAGCTAGTT 83567

Query Match   56.0%; Score 19.6; DB 1; Length 237326;
Best Local Similarity   73.5%; Pred. No. 10;
Matches 25; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

RESULT   2
V57926
ID   V57926 standard; DNA; 235033 BP.
AC   V57926;
DT   23-DEC-1998 (first entry)
DE   Hereditary haemochromatosis subregion from an unaffected individual.
DE   Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW

RESULT 1

LOCUS   AQ530999   462 bp   DNA   GSS   17-MAY-1999
DEFINITION RPCI-11-379C24.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-379C24, genomic survey sequence.
ACCESSION   AQ530999
NID   g4843042
VERSION   AQ530999.1 GI:4843042
KEYWORDS   GSS.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 462)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P., and
            Venter,J.C.
  TITLE   Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: RPCI-11-379C24.TJ
            Seq primer: T7
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
FEATURES   Location/Qualifiers
     source   1..462
                     /organism="Homo sapiens"
                     /db_xref="GDB:7645223"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-379C24"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Class: BAC ends.

ALIGNMENTS

c	#					Accession	Clone
	23	20	57.1	415	47	AI506876	v156e01.x
	24	20	57.1	874	80	AQ560640	HS_2066_B
	25	20	57.1	517	89	AQ757927	HS_5506_B
c	26	19.8	56.6	475	29	AA143686	z067c05.r
	27	19.8	56.6	431	42	AI115953	ue96e05.y
	28	19.8	56.6	432	42	AI115954	ue96e06.y
c	29	19.8	56.6	560	79	AQ312720	RPCI11-95
	30	19.8	56.6	405	29	AA162990	ms16g05.r
	31	19.6	56.0	424	69	AI603956	ms16g05.y
	32	19.6	56.0	1101	69	CNS00022	Drosophil
c	33	19.6	56.0	446	78	AQ404299	HS_5069_B
	34	19.6	56.0	492	90	AQ803131	HS_3158_B
	35	19.6	56.0	448	36	AA636172	vn15d03.r
	36	19.4	55.4	200	49	AI647109	vn15d03.y
	37	19.4	55.4	499	72	AQ180824	HS_3220_B
	38	19.4	55.4	522	81	AQ611367	HS_5088_B
c	39	19.4	55.4	574	90	AQ786710	HS_3189_A
c	40	19.4	55.4	226	20	T18141	070r63.czap
c	41	19.2	54.9	377	20	T18627	5c07a05-t7
	42	19.2	54.9	432	26	W39972	412 Mouse v
	43	19.2	54.9	445	28	AA119812	mn14c08.r
	44	19.2	54.9	433	28	AA120292	mn16c08.r
	45	19.2	54.9	579	34	AA511097	vh61g02.r

BASE COUNT   154 a   84 c   71 g   153 t
ORIGIN   "RPCI11 Human Male BAC Library"

Query Match   62.9%; Score 22; DB 80; Length 462;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy   1 CAGATCTAGATACTATGTTCTACACTCTTA 30
          |||| || ||| ||| |||||||||||||| |||
Db   183 CAGAACTCGATGCTATGTTCTACACCTTTA 212

RESULT 2

LOCUS   T79250   473 bp   mRNA   EST   15-MAR-1995
DEFINITION yd70c09.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:113584 5', mRNA sequence.
ACCESSION   T79250
NID   g697759
VERSION   T79250.1 GI:697759
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 473)
  AUTHORS   Hiller,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE   The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT   Other_ESTs: yd70c09.s1
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 688
            High quality sequence stops: 363 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 688   Std Error: 0.00
            Seq primer: M13R1
            High quality sequence stop: 363.
FEATURES   Location/Qualifiers
     source   1..473
                     /organism="Homo sapiens"
                     /db_xref="GDB:469201"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:113584"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAGAGCTCTTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT   135 a   73 c   76 g   185 t   4 others
ORIGIN

Query Match   59.4%; Score 20.8; DB 21; Length 473;

```
Best Local Similarity 78.1%; Conservative   0;   Pred. No. 75;
Matches      25;                 Mismatches   7;   Indels   0;   Gaps   0;

Qy     2 AGARCTAGAATACTARGTTCTACACTCTTAGGT 33
           ||||| |||||| ||||||||||||||| | ||
Db   313 ACAGTCACCTACTATGATCTACAATTTAGGT 344
```

RESULT 3

```
LOCUS       AA258261      411 bp    mRNA            EST        06-AUG-1997
            AA258261
DEFINITION  zr59g10.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:667746
            5', mRNA sequence.
ACCESSION   AA258261
NID         g1893423
VERSION     AA258261.1  GI:1893423
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,J., Wylie,T., Waterston,R. and Wilson,R.,
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393385.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 708   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source          1..411
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:667746"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NDHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223, and
                     340488-345479, and 484488-489479."
BASE COUNT      118 a     61 c     62 g    170 t
ORIGIN
```

```
Query Match 59.4%;  Score 20.8;  DB 30;  Length 411;
Best Local Similarity 78.1%;         Pred. No. 76;
Matches      25;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy     2 AGARCTAGAATACTARGTTCTACACTCTTACGT 33
           ||||| |||||| ||||||||||||||| ||
Db   256 ACAGTCACCTACTATGATCTACAAATTTAGGT 287
```

RESULT 4

```
LOCUS       AA465759      419 bp    mRNA            EST        11-JUN-1997
            AA465759
DEFINITION  vf07e06.r1 Knowles Solter mouse blastocyst B3 Mus musculus cDNA
            clone IMAGE:835042 5', mRNA sequence.
ACCESSION   AA465759
NID         g2191899
VERSION     AA465759.1  GI:2191899
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Morris,M.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:503248.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:495258
            High quality sequence stop: 396.
FEATURES             Location/Qualifiers
     source          1..419
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:835042"
                     /map="13q12"
                     /clone_lib="Knowles Solter mouse blastocyst B3"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts. Primer: SalI(dT):
                     5'-CGGTCGACGTCGACCGTTTTTTTTTTTTTTTT-3'; cDNAs were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies). Two different size selections: B1 (larger
                     inserts) and B3."
BASE COUNT      120 a     74 c     83 g    142 t
ORIGIN
```

```
Query Match 59.4%;  Score 20.8;  DB 34;  Length 419;
Best Local Similarity 78.1%;         Pred. No. 76;
Matches      25;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy     2 AGAICTAGAATACTARGTTCTACACTCTTACGT 33
           |||||||||| |  || || |||||||||| ||
Db    61 AGATCTAGAATGGTTGCTCCTTTACACTCTTAAGT 92
```

RESULT 5

```
LOCUS       AA669879      442 bp    mRNA            EST        12-MAY-1999
AA669879/c  AA669879
DEFINITION  ag42d01.s1 J1a bone marrow stroma Homo sapiens cDNA clone
            IMAGE:1119457 3', mRNA sequence.
ACCESSION   AA669879
NID         g2631378
```

```
VERSION     AA669879.1    GI:2631378
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 442)
AUTHORS     Jia,L., Robey,P., Young,M., Schuler,G., Powell,J., Yang,L.,
            Lennon,G., Hillier,L., Allen,M., Bowles,L., Geisel,S., Kucaba,T.,
            Marra,M., Martin,J., Steptoe,M., Tan,F., Theising,B., Bowers,Y.,
            Wylie,T., Waterston,R., Wilson,R. and Francomano,C.
TITLE       WashU-MGB/NHGRI EST Project
JOURNAL     Unpublished (1997)
COMMENT     WashU-MGB/NHGRI EST Project
            On Nov 6, 1997 this sequence version replaced gi:933403.
            Contact: Wilson RK / Jia L
            WashU-MGB/NHGRI EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 400.
FEATURES             Location/Qualifiers
     source          1..442
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1119457"
                     /clone_lib="Jia bone marrow stroma"
                     /sex="mixed"
                     /tissue_type="bone marrow stroma"
                     /dev_stage="mixed"
                     /lab_host="XL1-Blue MRF'/SOLR"
                     /note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
                     mRNA made from human bone marrow stroma, cDNA made by
                     oligo-dT priming. Directionally cloned. Size-selected for
                     average insert size >0.5 kb. Library supplied by Dr. Libin
                     Jia (NHGRI)."
BASE COUNT      181 a     64 c     66 g    131 t
ORIGIN

Qy          2 AGATCTAGAATACATATGTTCTACACTCTTACGT 33
              |||||| |||||||| ||||||| || |||||||
Db        161 ACATCTACCTACTATGATCTACAATTTTAGGT 130
```

```
RESULT  6
AA806372/c
                                                                Query Match     59.4%;  Score 20.8;  DB 37;  Length 442;
                                                                Best Local Similarity   78.1%;
                                                                Matches   25;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

LOCUS       AA806372        326 bp    mRNA            EST       25-MAR-1998
DEFINITION  oc22e08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1350470 3',
            mRNA sequence.
ACCESSION   AA806372
NID         g2875122
VERSION     AA806372.1    GI:2875122
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 326)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900811.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 902   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 279.
FEATURES             Location/Qualifiers
     source          1..326
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1350470"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGCGCGCCGCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      107 a     43 c     49 g    127 t
ORIGIN

Qy          2 AGARCTAGAATACATANGTTCTACACTCTTACGT 33
              |||||| ||||||||| |||||| |||||||| ||
Db        321 ACATCTACCTACTATGATCTACAATTTTAGGT 290
```

```
RESULT  7
AI378554/c
                                                                Query Match     59.4%;  Score 20.8;  DB 38;  Length 326;
                                                                Best Local Similarity   78.1%;
                                                                Matches   25;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

LOCUS       AI378554        485 bp    mRNA            EST       18-MAR-1999
DEFINITION  tc72a02.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:2070122
            3', mRNA sequence.
ACCESSION   AI378554
NID         g4188407
VERSION     AI378554.1    GI:4188407
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 485)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Oct 8, 1998 this sequence version replaced gi:3726019.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
```

FEATURES
source
```
                    Email: Robert_Strausberg@nih.gov
                    This clone is available royalty-free through LLNL; contact the
                    IMAGE Consortium (info@image.llnl.gov) for further information.
                    Insert Length: 449   Std Error: 0.00
                    Seq primer: -40UP from Gibco
                    High quality sequence stop: 443.
     1. .485
                    Location/Qualifiers
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:2070122"
                    /clone_lib="Soares_NhHMPu_S1"
                    /tissue_type="Pooled human melanocyte, fetal heart, and
                    pregnant uterus"
                    /lab_host="DH10B"
                    /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                    (Pharmacia) with a modified polylinker; Site_1: Not I;
                    Site_2: Eco RI; Equal amounts of plasmid DNA from three
                    normalized libraries (melanocyte 2NbHM, pregnant uterus
                    NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                    were made in vitro. Following HAP purification, this DNA
                    was used as tracer in a subtractive hybridization
                    reaction. The driver was PCR-amplified cDNAs from pools of
                    5,000 clones made from the same 3 libraries. The pools
                    consisted of I.M.A.G.E. clones 260232-265223,
                    340488-345479, and 484488-489479."
```

BASE COUNT   194 a   75 c   72 g   144 t
ORIGIN

```
Qy    2 AGATCTAGATACTAATGTTCTACACTCTTACGT 33
        ||||||| |||||||||||||||||||| ||||
Db  163 ACATCTACCTACTATGACTCTACAAATTTAGGT 132
```

RESULT 8
AI754738/c
```
LOCUS       AI754738      527 bp     mRNA        EST       22-JUN-1999
DEFINITION  cr29e04.x1 Jia bone marrow stroma Homo sapiens cDNA clone
            HBMSC.cr29e04 3', mRNA sequence.
ACCESSION   AI754738
NID         g5133002
VERSION     AI754738.1   GI:5133002
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
AUTHORS     Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,
            Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M.,
            Robey,P.G., Hotchkiss,R.N. and Francomano,C.A.
TITLE       SGAP: The Skeletal Genome Anatomy Project
JOURNAL     Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135954.

            Contact: Libin Jia
            Medical Genetics Branch
            National Human Genome Research Institute
            10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
            Tel: 301-402-4877
            Fax: 301-496-7157
            Email: libin@helix.nih.gov
            DNA Sequencing and analyses by National Institutes of Health
            Intramural Sequencing Center (NISC).
            Plate: 29   row: e   column: 04
            Seq primer: -21M13 forward primer (ABI).
     1. .527
            Location/Qualifiers
```

FEATURES
source

Query Match   59.4%;   Score 20.8;   DB 45;   Length 485;
Best Local Similarity   78.1%;   Pred. No. 74;
Matches 25; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

```
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="HBMSC_cr29e04"
                    /clone_lib="Jia bone marrow stroma"
                    /sex="mixed"
                    /tissue_type="bone marrow stroma"
                    /dev_stage="mixed"
                    /lab_host="XL1-Blue MRF'/SOLR"
                    /note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
                    mRNA made from human bone marrow stroma, cDNA made by
                    oligo-dT priming. Directionally cloned. Size-selected for
                    average insert size >0.5 kb. Library supplied by Dr. Libin
                    Jia (NHGRI)."
```

BASE COUNT   195 a   74 c   77 g   180 t   1 others
ORIGIN

```
Qy    2 AGATCTAGATACTAATGTTCTACACTCTTACGT 33
        |||||| ||||||||||||| |||||| ||| ||
Db  200 ACATCTACCTACTAGATCTACAAATTTAGGT 169
```

RESULT 9
AI769109/c
```
LOCUS       AI769109      318 bp     mRNA        EST       28-JUN-1999
DEFINITION  wg32h02.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:2366835 3', mRNA sequence.
ACCESSION   AI769109
NID         g5235618
VERSION     AI769109.1   GI:5235618
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 318)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135517.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 272.
     1. .318
            Location/Qualifiers
```

FEATURES
source
```
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:2366835"
                    /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                    /lab_host="DH10B"
                    /map="17q"
                    /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                    a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                    Equal amounts of plasmid DNA from five normalized
                    libraries were mixed, and ss circles were made in vitro.
                    Following HAP purification, this DNA was used as tracer in
                    a subtractive hybridization reaction. The driver was
                    PCR-amplified cDNAs from pools of 5,000 clones made from
                    the same 5 libraries. The pools consisted of the following
                    libraries and cloneIDs: Soares NbHSF pool 1:
                    309384-310919, 323208-325895 Soares Nb2HP pool 1;
                    145032-147335, 147720-148103, 148872-149255, 15002-
                    150407, 151176-152227 Soares Nb2HF8-9w pool 1:
```

Query Match   59.4%;   Score 20.8;   DB 51;   Length 527;
Best Local Similarity   78.1%;   Pred. No. 73;
Matches 25; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

BASE COUNT
ORIGIN

758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NbHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
105 a   45 c   49 g   119 t

Qy   2 AGATCTAGAATACTATAGTTCTACACTCTTACGT 33
            | ||||| |||||||| ||||||| ||||| ||
Db   304 ACATCTACCTACTATGATCTACACAAATTTAGGT 273

Query Match   59.4%; Score 20.8; DB 59; Length 318;
Best Local Similarity 78.1%; Pred. No. 79;
Matches 25; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

RESULT 10
AQ002323/c
LOCUS   AQ002323   489 bp   DNA   GSS   25-JUN-1998
DEFINITION CIT-HSP-2285G6.TF CIT-HSP Homo sapiens genomic clone 2285G6,
            genomic survey sequence.
ACCESSION   AQ002323
NID   g3029527
VERSION   AQ002323.1 GI:3029527
KEYWORDS   GSS.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 489)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE   Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: CIT-HSP-2285G6.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..489
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2285G6"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site.1: HindIII; Site_2:
                     HindIII"
BASE COUNT   199 a   73 c   68 g   149 t
ORIGIN

Qy   2 AGATCTAGAATACTATAGTTCTACACTCTTACGT 33
            | ||||| |||||||| ||||||| ||||| ||
Db   72 ACATCTACCTACTATGATCTACACAAATTTAGGT 41

Query Match   59.4%; Score 20.8; DB 70; Length 489;
Best Local Similarity 78.1%; Pred. No. 74;
Matches 25; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

RESULT 11
W15735/c
LOCUS   W15735   1311 bp   mRNA   EST   10-SEP-1996
DEFINITION mb53g03.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:333172 5', mRNA sequence.
ACCESSION   W15735
NID   g1290121
VERSION   W15735.1 GI:1290121
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1311)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE   The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:214572
            Seq primer: ETprimer
            High quality sequence stop: 462.
FEATURES   Location/Qualifiers
     source   1..1311
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:333172"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site.1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGACGGCGCGCATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT   221 a   216 c   559 g   272 t   43 others
ORIGIN

Qy   1 CAGATCTAGAATACTATAGTTCTACACTCTTACGTGC 35
            | ||||||| || |||||||||||| || ||||| ||
Db   926 CGGACCTAAAAACTATGTTGTAAACGCTTACAGGC 892

Query Match   58.9%; Score 20.6; DB 25; Length 1311;
Best Local Similarity 74.3%; Pred. No. 78;
Matches 26; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

RESULT 12
AI981758/c
LOCUS   AI981758   623 bp   mRNA   EST   15-SEP-1999
DEFINITION pat.pk0064.e2.f chicken activated T cell cDNA Gallus gallus cDNA
            clone pat.pk0064.e2.f 5' similar to probable cell cycle control
            protein crn, mRNA sequence.
ACCESSION   AI981758

NID   95884786
VERSION   AI981758.1 GI:5884786
KEYWORDS   EST.
SOURCE   chicken.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 623)
AUTHORS   Tirunagaru,V.G., Sofer,L. and Burnside,J.
TITLE   An expressed-sequence-tag database of activated chicken T cells:
            sequence analysis of 5596 clones
JOURNAL   Unpublished (1999)
COMMENT   On May 18, 1998 this sequence version replaced gi:3136799.
            Contact: Joan Burnside
            Molecular Endocrinology
            University of Delaware
            40 Townsend Hall, Newark, DE 19717, USA
            Tel: 302 831-1345
            Fax: 302 831-3651
            Email: Joan@Udel.Edu
            Seq primer: T7
FEATURES   Location/Qualifiers
source   1..623
            /organism="Gallus gallus"
            /db_xref="taxon:9031"
            /map="4; 4; 21q"
            /clone="pat.pk0064.e2.f"
            /clone_lib="chicken activated T cell cDNA"
            /sex="male"
            /cell_type="Con A-activated splenic T cell"
            /lab_host="E.coli TOP10 F'"
            /note="Vector: pcDNA3"
BASE COUNT   218 a   101 c   145 g   153 t   6 others
ORIGIN

Qy   1 CAGARCTAGATACTARGTTCTACACTCTTACGTGC 35
              |||||| |||||||||||||| |||||||||||||
Db   523 CAGARCCAGAGATGAATCTTCTTCCTTACGTTC 489

RESULT 13
AQ290000/c
LOCUS   AQ290000   582 bp   DNA   GSS   02-DEC-1998
DEFINITION nbxb0036M03f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0036M03f, genomic survey sequence.
ACCESSION   AQ290000
NID   g3951362
VERSION   AQ290000.1 GI:3951362
KEYWORDS   GSS.
SOURCE   Oryza sativa.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1 (bases 1 to 582)
AUTHORS   Yu,Y., Budiman,M.A., Boyd,C.A., Gaudette,B., Gayle,A.,
            Kingsbury,R., Phillips,R., Sasinowski,M., Dean,R.A. and Wing,R.A.
TITLE   A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL   Unpublished (1998)
COMMENT   Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu

Query Match   58.9%; Score 20.6; DB 63; Length 623;
Best Local Similarity   74.3%; Pred. No. 86;
Matches   26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

FEATURES   Location/Qualifiers
            Seq primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence stop: 483.
source   1..582
            /organism="Oryza sativa"
            /strain="Japonica"
            /cultivar="Nipponbare"
            /db_xref="taxon:4530"
            /clone="nbxb0036M03f"
            /clone_lib="CUGI Rice BAC Library"
            /tissue_type="leaf"
            /lab_host="E.coli DH10B"
            /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
            HindIII. Rice is one of two most popular grains in the
            world. Half of the world population especially those
            inhabiting highly populated areas of the humid tropics
            and subtropics, rely on rice as their primary source of
            carbohydrate. Monocotyledonous rice is a diploid plant
            (2n=24) with a haploid genome equivalent of 431 Mbp
            (Arumuganathan and Earle, 1991). The relatively small
            genome of rice, three times larger than that of
            Arabidopsis, makes it suitable for genomic studies. In
            order to facilitate positional cloning, physical mapping
            and genome sequencing of rice, we have constructed a BAC
            library from Oryza sativa, Nipponbare variety. The
            library contains 36,864 clones with an average insert size
            of 128.5 Kb providing 10.9 haploid genome equivalents.
            The deep coverage allows the isolation of a particular
            sequence with a probability of 99.9 %. Two high density
            filters, each containing 18,432 clones (doubly spotted),
            represent the whole library for colony screening."
BASE COUNT   169 a   95 c   131 g   187 t
ORIGIN

Qy   4 ATCTAGATACTARGTTCTACACTCTTA 30
              || |||| ||||||||||||| | |||||
Db   395 ARGTAGGTACTARGTTCTARTATCTTA 369

RESULT 14
AQ509564/c
LOCUS   AQ509564   545 bp   DNA   GSS   03-MAY-1999
DEFINITION nbxb0096H07r CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0096H07r, genomic survey sequence.
ACCESSION   AQ509564
NID   g4733647
VERSION   AQ509564.1 GI:4733647
KEYWORDS   GSS.
SOURCE   Oryza sativa.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 545)
AUTHORS   Yu,Y., Budiman,M.A., Boyd,C.A., Gaudette,B., Gayle,A.,
            Kingsbury,R., Phillips,R., Sasinowski,M., Dean,R.A. and Wing,R.A.
TITLE   A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL   Unpublished (1998)
COMMENT   Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG

Query Match   58.9%; Score 20.6; DB 77; Length 582;
Best Local Similarity   85.2%; Pred. No. 87;
Matches   23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

FT
FT
FT   /*tag= b
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PD   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 620-622; 327lpp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 4854 BP;   1615 A;   650 C;   986 G;   1480 T;

Query Match   49.1%;   Score 17.2;   DB 1;   Length 4854;
Best Local Similarity   71.0%;   Pred. No. 67;
Matches 22;   Conservative   0;   Mismatches   9;   Indels   0;   Gaps   0;

Qy   4 TAAGCTTGGAGTCACGTACGAGCAAGCTAGT 34
             |||||||| | ||||||| |||||| |
Db   2908 TGAGCTGTGGCTTATCGTACGANGAACGGATT 2938

```
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 272; 2245pp; Japanese.
CC   Claim 1: Page 272; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA), which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence 574 BP; 148 A; 109 C; 147 G; 141 T;

Query Match      50.9%;   Score 17.8;   DB 1;   Length 574;
Best Local Similarity  75.9%;   Pred. No. 26;
Matches  22;   Conservative  0;   Mismatches  7;   Indels  0;   Gaps  0;

Qy    1 TCTTAAGCTTGGAGTCACGTACGGAGCAAG 29
          ||| ||||| |||||||| | ||||||||
Db  221 TCTCAAGCCAGGAGTCAAAGAAGAGCAAG 249
```

```
RESULT   6
ID   T26895   standard; DNA; 1537 BP.
AC   T26895;
DT   15-AUG-1996 (first entry)
DE   Bacteriophage resistance Abi determinant of plasmid pND859.
KW   Bacteriophage resistance; phage resistance; lactic acid bacterium;
KW   starter culture; dairy; food processing; plasmid pND852; AbiJ;
KW   abortive infection mechanism; ss.
OS   Lactococcus lactis var. diacetylactis strain UK 12922.
FH   Key              Location/Qualifiers
FT   -35_signal       96..101
FT                    /tag= a
FT   -10_signal       120..125
FT                    /tag= b
FT   rbs              154..159
FT                    /tag= c
FT   cds              174..1022
FT                    /tag= d
PN   GB2294463-A.
PD   01-MAY-1996.
PF   13-OCT-1995; 020797.
PR   13-OCT-1994; AU-008764.
PA   (ARNO-) ARNOTT'S BISCUITS LTD.
PA   (BURN-) BURNS PHILP RES & DEV PTY LTD.
PA   (BURN-) BURNS PHILP & CO LTD.
PA   (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA   (GOOD-) GOODMAN FIELDER INGREDIENTS LTD.
PA   (MAUR) MAURI LAB PTY LTD.
PA   (UYNE-) UNIV NEW SOUTH WALES.
PI   Costello M, Deng Y, Dunn N, Harvey M, Heejeong I;
PI   Kennelly V, Su P;
DR   WPI; 96-202561/21.
DR   P-PSDB; R95450.
PT   New plasmids, and derived nucleic acid, that impart bacteriophage
PT   resistance - partic. to lactic acid bacteria used in food
PT   processing
CC   Claim 4; Fig 9; 53pp; English.
CC   A 2.9 kb HindIII fragment from pND859 of Lactococcus lactis var.
CC   diacetylactis UK 12922 includes a sequence (T26895) coding for an
CC   Abi (abortive infection mechanism) bacteriophage resistance
CC   determinant (R95450). Plasmid pND859 can be conjugatively
CC   transferred to lactococcal recipients where it confers insensitivity
CC   to phage infection and resistance to cadmium. It was isolated by
CC   filter-mate conjugation. pND859 or the AbiJ determinant can be used
CC   to impart, or to increase, the phage resistance of lactic acid
CC   bacteria used in food processing.
SQ   Sequence 1537 BP; 564 A; 207 C; 267 G; 499 T;

Query Match      50.3%;   Score 17.6;   DB 1;   Length 1537;
Best Local Similarity  71.9%;   Pred. No. 37;
Matches  23;   Conservative  0;   Mismatches  9;   Indels  0;   Gaps  0;

Qy    3 TTAAGCTTGGAGTCACGTACGGAGCAAGCTAGT 34
          || |  ||||||| | |||||| ||||| |||
Db  205 TAAAATTGGAGTAAGTTACGAGGGAGGTAGT 236
```

```
RESULT   7
ID   V84620/c
AC   V84620;
     V84620/c   standard; DNA; 3018 BP.
DT   01-MAR-1999 (first entry)
DE   Human secreted protein gene 19 clone HSAVU34.
KW   Human secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
PN   WO9854963-A2.
PD   10-DEC-1998.
PF   04-JUN-1998;                 U11422.
PF   18-DEC-1997;                 US-070923.
PR   06-JUN-1997;                 US-048877.
PR   06-JUN-1997;                 US-048881.
PR   06-JUN-1997;                 US-048884.
PR   05-SEP-1997;                 US-048893.
PR   06-JUN-1997;                 US-048896.
PR   06-JUN-1997;                 US-048899.
PR   06-JUN-1997;                 US-048915.
PR   06-JUN-1997;                 US-048935.
PR   06-JUN-1997;                 US-048949.
PR   06-JUN-1997;                 US-048964.
PR   06-JUN-1997;                 US-048972.
PR   06-JUN-1997;                 US-049020.
PR   06-JUN-1997;                 US-049375.
PR   05-SEP-1997;                 US-057628.
PR   05-SEP-1997;                 US-057635.
PR   05-SEP-1997;                 US-057644.
PR   05-SEP-1997;                 US-057647.
PR   05-SEP-1997;                 US-057650.
PR   05-SEP-1997;                 US-057661.
PR   05-SEP-1997;                 US-057667.
PR   05-SEP-1997;                 US-057761.
PR   05-SEP-1997;                 US-057764.
PR   05-SEP-1997;                 US-057770.
PR   05-SEP-1997;                 US-057775.
PR   05-SEP-1997;                 US-057778.
PR   06-JUN-1997;                 US-048875.
PR   06-JUN-1997;                 US-048878.
PR   06-JUN-1997;                 US-048882.
PR   06-JUN-1997;                 US-048885.
PR   06-JUN-1997;                 US-048894.
PR   06-JUN-1997;                 US-048897.
PR   06-JUN-1997;                 US-048900.
PR   06-JUN-1997;                 US-048916.
```

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: December 29, 1999, 04:26:25 ; Search time 727.51 Seconds
                                             (without alignments)
                                             164.915 Million cell updates/sec

Title: US-09-077-173A-3
Perfect score: 35
Sequence: 1 CAGARCTAGATAACTATATGTTCTACACTCTTACGTGC 35

Scoring table: IDENTITY_NUC

Searched: 4089388 seqs, 1713965092 residues

Database :   EST:*

Word size :   0

Number of hits that pass the threshold :   8178776

```
 1: em_est1:*          50: gb_est31:*
 2: em_est2:*          51: gb_est32:*
 3: em_est3:*          52: em_est20:*
 4: em_est4:*          53: em_est21:*
 5: em_est5:*          54: em_est22:*
 6: em_est6:*          55: em_est23:*
 7: em_est7:*          56: em_est24:*
 8: em_est8:*          57: em_est25:*
 9: em_est9:*          58: em_est26:*
10: em_est10:*         59: gb_est33:*
11: em_est11:*         60: gb_est34:*
12: em_est12:*         61: gb_est35:*
13: em_est13:*         62: gb_est36:*
14: em_est14:*         63: gb_est37:*
15: em_est15:*         64: gb_est38:*
16: em_est16:*         65: em_est27:*
17: em_est17:*         66: em_est28:*
18: em_est18:*         67: em_est29:*
19: em_est19:*         68: em_est30:*
20: em_est1:*          69: gb_gss1:*
21: em_est2:*          70: gb_gss2:*
22: gb_est3:*          71: gb_gss3:*
23: gb_est4:*          72: gb_gss4:*
24: gb_est5:*          73: em_gss1:*
25: gb_est6:*          74: em_gss2:*
26: gb_est7:*          75: em_gss3:*
27: gb_est8:*          76: em_gss4:*
28: gb_est9:*          77: gb_gss5:*
29: gb_est10:*         78: gb_gss6:*
30: gb_est11:*         79: gb_gss7:*
31: gb_est12:*         80: gb_gss8:*
32: gb_est13:*         81: gb_gss9:*
33: gb_est14:*         82: em_gss5:*
34: gb_est15:*         83: em_gss6:*
35: gb_est16:*         84: em_gss7:*
36: gb_est17:*         85: em_gss8:*
37: gb_est18:*         86: em_gss9:*
38: gb_est19:*         87: em_gss10:*
39: gb_est20:*         88: em_gss11:*
40: gb_est21:*         89: em_gss10:*
41: gb_est22:*         90: gb_gss11:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		22	62.9	462	80	AQ530999	AQ530999 RPCI-11-3
2		20.8	59.4	473	21	T79250	T79250 yd70c09.r1
3		20.8	59.4	411	30	AA258261	AA258261 zr59g10.r
4		20.8	59.4	419	34	AA465759	AA465759 vf07e06.r
5		20.8	59.4	442	37	AA669879	AA669879 ag42d01.s
6	c	20.8	59.4	326	38	AA806372	AA806372 oc22e08.s
7	c	20.8	59.4	485	45	AI378554	AI378554 tc72a02.x
8	c	20.8	59.4	527	51	AI754738	AI754738 cr29e04.x
9	c	20.8	59.4	318	59	AI769109	AI769109 wg32h02.x
10	c	20.8	59.4	489	70	AQ002323	AQ002323 CIT-HSP-2
11	c	20.6	58.9	1311	25	W15735	W15735 mb53g03.r1
12	c	20.6	58.9	623	63	AI981758	AI981758 pat.px006
13	c	20.6	58.9	582	77	AQ290000	AQ290000 nbxb0036M
14	c	20.6	58.9	545	79	AQ509564	AQ509564 nbxb0096H
15	c	20.6	58.9	364	81	AQ690006	AQ690006 nbxb0081G
16	c	20.6	58.9	207	90	AQ794814	AQ794814 nbxb0054G
17	c	20.4	58.3	431	71	AQ111301	AQ111301 CIT-HSP-2
18	c	20.2	57.7	439	22	R49688	R49688 yg64d11.s1
19		20.2	57.7	396	70	AQ010184	AQ010184 HS_2263_B
20	c	20.2	57.7	371	71	AQ104808	AQ104808 HS_2166_B
21	c	20.2	57.7	475	77	AQ296901	AQ296901 HS_3005_B
22	c	20.2	57.7	466	79	AQ493046	AQ493046 HS_5222_B

KW   diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTFI; BTF2; BTF3;
KW   BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW   type 1 sodium transport gene; ss.
OS   Homo sapiens.
PN   WO9814466-A1.
PD   09-APR-1998;
PD   30-SEP-1997; U17658.
PD   07-MAY-1997; US-852495.
PR   01-OCT-1996; US-724394.
PA   (PROG-) PROGENTOR INC.
PI   Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI   Tsuchihashi Z, Wolff RK;
DR   WPI; 98-240014/21.
PT   Hereditary haemochromatosis gene products - used to develop products
PT   for the diagnosis and treatment of hereditary disorders in iron
PS   metabolism
PS   Example 2; Fig 8; 209pp; English.
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. The present sequence
CC   represents a hereditary haemochromatosis subregion from an individual
CC   unaffected by hereditary haemochromatosis (HH). Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   presence of the HFE gene mutation in the genome of the individual. The
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes BTF genes, which are homologues of the milk
CC   protein butyrophilin (BT), and can be used in the production of agonists
CC   and antagonists of BT function. Also described are: (1) a RoRet gene
CC   which can be used to develop products for the study, diagnosis and
CC   treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC   which are homologues of a type 1 sodium transport gene, and can
CC   similarly be used for hypophosphatemia.
SQ   Sequence 235033 BP;   68786 A;   48466 C;   49441 G;   68340 T;

Qy   2 CTTAAGCTTGGAGTCACGTACGAGCGAAGCTAGTT 35
        |||||||||| || ||||| ||||||||||||||
Db 83581 CTAAAGCTTGGGTGTCATTAAGGAGTGTAGCTAGTT 83614

Query Match   56.0%; Score 19.6; DB 1; Length 235033;
Best Local Similarity   73.5%;
Matches   25; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

RESULT   3
ID   Q47841 standard; mRNA to cDNA; 1927 BP.
AC   Q47841;
DT   28-MAR-1994   (first entry)
DE   Flavonoid-3',5'-hydroxylase gene.
KW   Flavonoid-3',5'-hydroxylase; transformation; plants; petunia; rose;
KW   tobacco; pigment alteration; blue; SSP; single specific primer; ss.
OS   Campanula medium.
FH   Key   Location/Qualifiers
FT   cds   180..1751
FT   /*tag= a
FT   /product= Flavonoid-3',5'-hydroxylase
PN   WO9318155-A.
PD   16-SEP-1993;
PF   20-NOV-1992; J01520.
PF   02-MAR-1992; JP-044963.
PA   (KYOW) KYOWA HAKKO KOGYO CO LTD.
PI   Kikuchi Y, Kiyokawa S, Ohbayashi M, Okinaka Y, Shimada R;
PI   Shimada Y;
WPI; 93-303469/38.
DR   P-PSDB; R40868.
DR   Gene coding for flavonoid-3',5'-hydroxylase of petunia petals -
PT   used to transform plants e.g. petunia, rose or tobacco to give
PT   bluer flower colour and altered pigment pattern

PS   Claim 1; Page 71-75; 82pp; Japanese.
CC   Insertion of the sequences (Q478401-42) into plants such as rose,
CC   petunia, tobacco and carnation, using a suitable vector such as
CC   agrobacterium, give transformed plants which express the gene,
CC   resulting in petals with a bluer colour than normal, and/or
CC   pigmentation patterns which do not occur naturally. The sequences
CC   were amplified using primers (Q47843-70). Related single specific
CC   primers using a gene sequence coding for the haem-binding region of
CC   cytochrome P450 are shown in (Q47871-Q47903).
SQ   Sequence 1927 BP;   582 A;   399 C;   396 G;   550 T;

Query Match   51.4%; Score 18; DB 1; Length 1927;
Best Local Similarity   70.6%;
Matches   24; Conservative   0; Mismatches   10; Indels   0; Gaps   0;

Qy   1 TCTTAAGCTTGGAGTCACGTACGAGCGAAGCTAGT 34
        || || |||||| || | | |||||||||| |||
Db 656 TCATATGCTCAAAGAAAATGTACGAGCAATCGAGT 689

RESULT   4
ID   Q90096 standard; cDNA; 2600 BP.
AC   Q90096;
DT   03-NOV-1995   (first entry)
DE   Mouse kappa-3 opioid receptor.
KW   Kappa-3 opioid receptor; analgesia; ss.
OS   Mus sp.
FH   Key   Location/Qualifiers
FT   cds   299..1402
FT   /*tag= a
PN   WO9512616-A.
PD   11-MAY-1995;
PD   03-NOV-1994; U12728.
PF   05-NOV-1993; US-147949.
PA   (SLOK) SLOAN KETTERING INST CANCER RES.
PI   Pan Y, Pasternak GW;
DR   WPI; 95-193814/25.
DR   P-PSDB; R74298.
PT   Nucleic acid molecule(s) encoding a kappa-3 opioid receptor, and
PT   antibody against the receptor - used to detect the receptor, and to
PT   image cell membrane-bound receptor in a subject
PS   Disclosure; Fig.1; 68pp; English.
CC   Degenerate primers based on conserved sequences of the mouse delta
CC   opioid receptor were used in PCR to amplify mouse kappa-3 opioid
CC   receptor cDNA from a mouse brain lambda ZAP cDNA library.
SQ   Sequence 2600 BP;   536 A;   680 C;   663 G;   721 T;

Qy   2 CTTAAGCTTGGAGTCACGTACGAGCA 27
        |||||||||||||||| | | |||||
Db 1778 CTTAAGCTTGGAGTTGCCTTGGGAGCA 1803

Query Match   51.4%; Score 18; DB 1; Length 2600;
Best Local Similarity   80.8%;
Matches   21; Conservative   0; Mismatches   26; Indels   5; Gaps   0;

RESULT   5
ID   T19014 standard; cDNA to mRNA; 574 BP.
AC   T19014;
DT   14-JUN-1996   (first entry)
DE   Human gene signature HUMGS00015.
KW   Human gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.

CC   or thymus, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
CC   The present sequence represents a gene encoding a human secreted protein
CC   (see descriptor line for gene number and clone identification).
SQ   Sequence 3018 BP; 841 A; 635 C; 678 G; 863 T;

Qy   3 TTAAGCTTGGAGTCACGTACGGAGCAAGCTAGT 34
       ||| || || |||||| |||| |||| ||||| ||
Db   1925 TAAAATCTGGGAGCCAGGTACGTGCAAAGTGGT 1894

RESULT 8

V84429/c

ID   V84429 standard; DNA; 3735 BP.
AC   V84429;
DT   01-MAR-1999 (first entry)
DE   Human secreted protein gene 19 clone HSAVU34.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens
PN   WO9854963-A2.
PD   10-DEC-1998.
PF   04-JUN-1998; U11422.
PR   18-DEC-1997; US-070923.
PR   06-JUN-1997; US-048915.
PR   06-JUN-1997; US-048949.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048877.
PR   06-JUN-1997; US-048881.
PR   06-JUN-1997; US-048884.
PR   06-JUN-1997; US-048893.
PR   06-JUN-1997; US-049375.
PR   05-SEP-1997; US-057628.
PR   05-SEP-1997; US-057635.
PR   05-SEP-1997; US-057644.
PR   05-SEP-1997; US-057647.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057661.
PR   05-SEP-1997; US-057667.
PR   05-SEP-1997; US-057761.
PR   05-SEP-1997; US-057764.
PR   05-SEP-1997; US-057770.
PR   05-SEP-1997; US-057775.
PR   05-SEP-1997; US-057778.
PR   06-JUN-1997; US-048875.
PR   06-JUN-1997; US-048878.
PR   06-JUN-1997; US-048882.
PR   06-JUN-1997; US-048885.
PR   06-JUN-1997; US-048894.
PR   06-JUN-1997; US-048897.
PR   06-JUN-1997; US-048900.
PR   06-JUN-1997; US-048907.
PR   06-JUN-1997; US-048916.
PR   06-JUN-1997; US-048960.
PR   06-JUN-1997; US-048962.
PR   06-JUN-1997; US-048970.
PR   06-JUN-1997; US-048974.
PR   06-JUN-1997; US-049373.
PR   05-SEP-1997; US-057584.
PR   05-SEP-1997; US-057629.

Query Match 50.3%; Score 17.6; DB 1; Length 3018;
Best Local Similarity 71.9%; Pred. No. 41;
Matches 23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PI   Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PI   Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI   Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI   Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR   WPI; 99-059865/05.
DR   P-PSDB; W88743.

PT   New isolated human genes and the secreted polypeptides they encode -
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
Ps   Claim 4; Page 472-473; 772pp; English.
CC   The invention relates to nucleic acid sequences (V84411 to V84633)
CC   encoding human secreted proteins (W88534 to W88756). The secreted protein
CC   gene sequences are deposited with the ATCC under deposit numbers ATCC
CC   97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC   209011, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC   cells comprising recombinant vectors containing the nucleic acid
CC   sequences are used for the recombinant production of the secreted
CC   proteins. The polynucleotide and amino acid sequences are useful for are
CC   useful for gene therapy. Pathological conditions can be also diagnosed
CC   by determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the polynucleotides, based on
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, neurodegenerative
CC   disorders, developmental abnormalities and foetal deficiencies, blood
CC   disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC   diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC   ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC   restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC   such as osteoporosis, arthritis or malignancies, diseases of testes, lung

PR   06-JUN-1997; US-048962.
PR   06-JUN-1997; US-048970.
PR   06-JUN-1997; US-048974.
PR   06-JUN-1997; US-049373.
PR   06-JUN-1997; US-057584.
PR   05-SEP-1997; US-057629.
PR   05-SEP-1997; US-057642.
PR   05-SEP-1997; US-057645.
PR   05-SEP-1997; US-057648.
PR   05-SEP-1997; US-057651.
PR   05-SEP-1997; US-057662.
PR   05-SEP-1997; US-057668.
PR   05-SEP-1997; US-057762.
PR   05-SEP-1997; US-057765.
PR   05-SEP-1997; US-057771.
PR   05-SEP-1997; US-057776.
PR   05-SEP-1997; US-057777.
PR   06-JUN-1997; US-048876.
PR   06-JUN-1997; US-048880.
PR   05-SEP-1997; US-057649.
PR   05-SEP-1997; US-057654.
PR   05-SEP-1997; US-057660.
PR   05-SEP-1997; US-057763.
PR   05-SEP-1997; US-057769.
PR   05-SEP-1997; US-057774.

PR   05-SEP-1997; US-057642.
PR   05-SEP-1997; US-057645.
PR   05-SEP-1997; US-057648.
PR   05-SEP-1997; US-057651.
PR   05-SEP-1997; US-057662.
PR   05-SEP-1997; US-057668.
PR   05-SEP-1997; US-057762.
PR   05-SEP-1997; US-057765.
PR   05-SEP-1997; US-057771.
PR   05-SEP-1997; US-057776.
PR   05-SEP-1997; US-049019.
PR   06-JUN-1997; US-049374.
PR   06-JUN-1997; US-048876.
PR   06-JUN-1997; US-048880.
PR   06-JUN-1997; US-048883.
PR   06-JUN-1997; US-048892.
PR   06-JUN-1997; US-048895.
PR   06-JUN-1997; US-048898.
PR   06-JUN-1997; US-048901.
PR   06-JUN-1997; US-048917.
PR   06-JUN-1997; US-048961.
PR   06-JUN-1997; US-048971.
PR   05-SEP-1997; US-057760.
PR   05-SEP-1997; US-057763.
PR   05-SEP-1997; US-057769.
PR   05-SEP-1997; US-057774.
PR   05-SEP-1997; US-057777.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PA   Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PA   Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI   Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI   Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR   WPI; 99-058865/05.
DR   P-PSDB; W88552, W88760, W88761.
DR   P-PSDB; W88552, W88760, W88761.
PT   New isolated human genes and the secreted polypeptides they encode -
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Claim 4; Page 287-289; 772pp; English.
CC   The invention relates to nucleic acid sequences (V84411 to V84633)
CC   encoding human secreted proteins (W88534 to W88756) The secreted protein
CC   gene sequences are deposited with the ATCC under deposit numbers ATCC
CC   209079, 209074, 209075, 209976, 209977, 209007, 209008, 209009, 209010,
CC   209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC   cells comprising recombinant vectors containing the nucleic acid
CC   sequences are used for the recombinant production of the secreted
CC   proteins. The polynucleotide and amino acid sequences are useful for are
CC   useful for preventing, treating or ameliorating medical conditions e.g.
CC   by determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the polynucleotides, based on
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, neurodegenerative
CC   disorders, developmental abnormalities and foetal deficiencies, blood
CC   disorders, tumours, leukemias diseases of the immune system, autoimmune
CC   diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC   ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC   restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC   such as osteoporsis, arthritis or malignancies, diseases of testes, lung
CC   or thymus, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners. The
CC   present sequence represents a gene encoding a human secreted protein
CC   (see descriptor line for gene number and clone identification).
SQ   Sequence 3735 BP; 1076 A; 773 C; 814 G; 1067 T;

Qy   3 TTAAGCTTGGAGTGCACGTACGAGCAAGCTAGT 34
            | ||| || ||| || ||||||| | |||||| | ||
Db   2962 TAAARCTGGGAGCCRGGTACGTGCRAAAGTGGT 2931

Query Match   50.38; Score 17.6; DB 1; Length 3735;
Best Local Similarity   71.9%; Pred. No. 42;
Matches   23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

RESULT   9
ID   V74942   standard; DNA; 4005 BP.
AC   V74942;
DT   16-MAR-1999 (first entry)
DE   Staphylococcus aureus contig SEQ ID #631.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
FT   misc_feature   661..720
FT   /*tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   2461..2520
FT   /*tag= b
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1557-1559; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 4005 BP; 1047 A; 716 C; 585 G; 1537 T;

Qy   4 TAAGCTTGGAGTGCACGTACGAGCAAGCTAGTT 35
            |||| ||| ||| |||| ||| ||| ||| || || ||
Query Match   50.38; Score 17.6; DB 1; Length 4005;
Best Local Similarity   71.9%; Pred. No. 43;
Matches   23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db　3504　TAAGTTTGGAATCAAGTATGACCCATCTTCTT 3535

RESULT 10

ID　T47698 standard; DNA; 2078 BP.
AC　T47698;
DT　12-AUG-1997 (first entry)
DE　Banana polyphenol oxidase gene (clone BANPO1).
KW　primer; degenerate; PCR; polymerase chain reaction; amplify; banana;
KW　lettuce; polyphenol oxidase; PPO; conserved region; copper binding site;
KW　CuA; CuB; GEN3; GEN8; antisense construct; decrease; prevent; browning;
OS　Musa sapientum; ss.
OS　Musa sapientum.
FH
FT　Key　　　Location/Qualifiers
FT　cds　　　2..1825
FT　　　　　/*tag= a
FT　　　　　/note= "no start codon is given and the reading frame is
FT　　　　　interupted by a stop codon at positions 38-40"
PN　WO9637617-A1.
PD　28-NOV-1996.
PF　22-MAY-1996; 96AU00310.
PR　23-MAY-1995; AU-003098.
PR　26-SEP-1995; AU-005600.
PA　(CSIR) COMMONWEALTH SCI & IND RES ORG.
PI　Robinson SP;
WPI; 97-033994/03.
DR　P-PSDB; W09422.
PT　DNA encoding banana or lettuce polyphenol oxidase - useful in
PT　preventing browning in plants.
PS　Claim 14; Fig 2; 33pp; English.
CC　The present sequence, BANPO1 encodes banana polyphenol oxidase (PPO).
CC　It was derived by combining the sequences from various clones (BPO3,
CC　BPO17 and BPO26) isolated using degenerate (T47685-87) and specific
CC　primers (T47692-94). A number of clones were obtained from the banana
CC　fruit cDNA using primers B25 and GEN7 (T47695-96). Most of these clones
CC　were identical to BANPO1 but one clone, designated BANPO1, was found
CC　to be distinctly different (see T47699). The primers were designed based
CC　on known plant PPO DNA sequences in the conserved regions of the gene
CC　which encode the copper binding sites, CuA and CuB. Antisense banana or
CC　lettuce PPO constructs can be used to decrease the level of PPO activity
CC　in a banana or lettuce plant. Decreasing the level of PPO is useful in
CC　preventing browning in plants.
SQ　Sequence 2078 BP; 489 A; 629 C; 561 G; 399 T;

Qy　5　AAGCTTGGAGTCACGTACGAGCAAGCT 31
　　　|||||| | |||| | |||||||||||| |
Db　1262　AAGCTTCGGTACACGTTCGAGCAAGTT 1288

Query Match　49.7%; Score 17.4; DB 1; Length 2078;
Best Local Similarity 77.8%; Pred. No. 48;
Matches　21; Conservative　0; Mismatches　6; Indels　0; Gaps　0;

RESULT 11

ID　Q40945 standard; DNA; 5433 BP.
AC　Q40945;
DT　08-SEP-1993 (first entry)
DE　Human DNA polymerase alpha catalytic DNA sequence (corrected).
KW　DNA polymerase; alpha catalytic polypeptide; hybridisation; ss;
KW　template-dependent enzymatic nucleic acid synthesis; frameshift.
OS　Homo sapiens.
FH
FT　Key　　　Location/Qualifiers
FT　cds　　　17..4405
FT　　　　　/*tag= a
FT　variation　1523..1534
FT　　　　　/*tag= b
FT　　　　　/note= "mutation from known sequence"
FT　variation　2525..2527
FT　　　　　/*tag= c
FT　　　　　/note= "mutation from known sequence"

PN　WO9310139-A.
PD　27-MAY-1993.
PF　12-NOV-1992; U09457.
PR　15-NOV-1991; US-792600.
PA　(STRD) UNIV LELAND STANFORD JUNIOR.
PA　Copeland WC, Wang TSF;
WPI; 93-182482/22.
P-PSDB; R37508.
PT　Purified nucleic acid for enzymatic synthesis of nucleic acid -
PT　encodes human polymerase alpha catalytic polypeptide, and is
PT　contained in baculovirus vector for enhanced prodn. in insect
PT　cells
PS　Claim 2; Fig 3; 84pp; English.
CC　Human DNA polymerase alpha catalytic polypeptides were isolated from
CC　a human epidermoid carcinoma cell line, and seven peptides, T9, T19,
CC　T23, T24, T25, T64 and T265, were sequenced. Each peptide sequence
CC　was used to design anti-sense probes which were used to probe human
CC　KB cDNA libraries to recover positive colonies contg. fragments
CC　encoding DNA polymerase polypeptides. One cDNA clone obtd. was
CC　used to screen a human pre-B cell cDNA library in lambda gt10 to
CC　obtain a full length clone. Initial attempts to functionally express
CC　the clone for human DNA polymerase alpha catalytic subunit resulted
CC　in truncated translation prods. Resequencing of five overlapping
CC　clones in conjunction with in vitro translation analysis revealed
CC　two frame shift mutations and two missense mutations in the two
CC　previously isolated cDNA clones, E1-19 and E1-12, that contain the 5'
CC　end of the cDNA sequence. The sequence was mutated to correct the
CC　variations and a full length ORF of 1462 amino acids produced.
CC　See also Q40921-49.
SQ　Sequence 5433 BP; 1672 A; 1052 C; 1272 G; 1437 T;

Qy　5　AAGCTTGGAGTCACGTACGGAGCAAGCTAGT 34
　　　|||||| ||| ||||| || |||||||| |
Db　4748　AAGCCTGGGTCACTTACATTCCAGCTACT 4719

Query Match　49.1%; Score 17.2; DB 1; Length 5433;
Best Local Similarity 73.3%; Pred. No. 68;
Matches　22; Conservative　0; Mismatches　8; Indels　0; Gaps　0;

RESULT 12

ID　T13635 standard; DNA; 133894 BP.
AC　T13635;
DT　03-SEP-1996 (first entry)
DE　AcNPV genomic DNA clone 6.
KW　Autographa californica nuclear polyhedrosis virus clone 6;
KW　disruption; non-essential gene; heterologous protein production;
KW　expression vector; baculovirus; ss.
OS　Autographa californica nuclear polyhedrosis virus clone 6.
KW
PN　WO9601320-A2.
PD　18-JAN-1996.
PF　30-JUN-1995; IB0578.
PR　04-JUL-1994; GB-013420.
PA　(NAPU) NATURAL ENVIRONMENT RES COUNCIL.
PA　Ayres M, Bishop D, Possee R;
WPI; 96-087670/09.
DR　GENBANK; L22858.
PT　Autographa californica nuclear polyhedrosis virus complete genome
PT　sequence - useful in the prodn. of vectors for enhanced
PT　heterologous protein expression, such as interleukin(s),
PT　interferon(s) and neurotoxin(s)
PS　Disclosure; Page 90-186; 122pp; English.
CC　The complete nucleotide sequence of the genome of clone 6 of the
CC　baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)
CC　has been determined. The sequence is taken from the Genbank record
CC　L22858. The patent specification claims a polynucleotide selected from
CC　open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,
CC　62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,
CC　140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
CC　the patentees. See T13636-731. Expression vectors contg. the complete
CC　genomic sequence of AcNPV, with the exception that at least one non-

CC   essential ORF is disrupted or replaced are useful for the synthesis of
CC   heterologous proteins.
SQ   Sequence 133894 BP;   39195 A;   27151 C;   27347 G;   40201 T;

Qy   1 TCTTAAGCTTGGAGTCAGTCACGTACGAGCAAGC 30
        | ||||| |||| ||| ||| |||||||||| ||| |
Db 10003 TGTTAAGTTTGGCGTCGAGTTCGTGCAAAC 10032

Query Match   49.1%; Score 17.2; DB 1; Length 133894;
Best Local Similarity   73.3%; Pred. No. 1.1e+02;
Matches   22; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

RESULT 13
ID   T13636/c
AC   T13636;
     T13636;
DT   06-SEP-1996 (first entry)
DE   AcNPV ORF 13, residues 10621-9683.
KW   Autographa californica nuclear polyhedrosis virus clone 6;
KW   disruption; non-essential gene; heterologous protein production;
KW   expression vector; baculovirus; ss.
OS   Autographa californica nuclear polyhedrosis virus clone 6.
FH   Key   Location/Qualifiers
cds   1..984
FT   /tag= a
FT   /number= 13
FT   /note= "corresponds to AcNPV nucleotides
FT   complement (10621-9683)"

PN   WO9601320-A2.
PD   18-JAN-1996.
PF   30-JUN-1995; IB0578.
PR   04-JUL-1994; GB-013420.
PA   (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
PI   Ayres M, Bishop D, Possee R;
DR   WPI; 96-087670/09.
DR   GENBANK; L22858.
PT   Autographa californica nuclear polyhedrosis virus complete genome
PT   sequence - useful in the prodn. of vectors for enhanced
PT   heterologous protein expression, such as interleukin(s),
PT   interferon(s) and neurotoxin(s)
PS   Claim 1; Page 90-186; 122pp; English.
PT   T13636-731 show open reading frames 13, 20, 22-30, 32, 38, 41-46, 50-60,
PT   62-63, 66, 81-87, 91-92, 96-98, 101-103, 106-127, 129-130, 140-
PT   146, 148-150, 152 and 154 from a total of 154 ORFs identified in the
CC   gene is numbered according to its position in the virus genome beginning
CC   at the left end of the linear map, and irrespective of its orientation.
CC   The direction of transcription is relative to that of the polyhedrin
CC   gene. Of the 154 ORFs identified it was found that some of the ORFs (ORFS
CC   27, 30, 32, 71, 86, 123, 126 and 127) are dispensable for virus
CC   replication in cell culture or insect larvae. These genes can be deleted
CC   from the genome to: (a) provide additional sites for inserting single or
CC   multiple copies of foreign genes; and (b) to reduce the size of the virus
CC   genome. The present sequence ORF 13 is on the complementary strand
CC   relative to the polyhedrin gene.
SQ   Sequence   984 BP;   314 A;   211 C;   227 G;   232 T;

Qy   1 TCTTAAGCTTGGAGTCAGTCACGTACGAGCAAGC 30
        | ||||| |||| ||| ||| ||||| ||| ||| |
Db 619 TGTTAAGTTTGGCGTCGAGTTCGTGCAAAC 590

Query Match   49.1%; Score 17.2; DB 1; Length 984;
Best Local Similarity   73.3%; Pred. No. 52;
Matches   22; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

RESULT 14
ID   V03349
AC   V03349;
DE   V03349 standard; DNA; 1679 BP.

DT   17-JUN-1998 (first entry)
DE   Homo sapiens interleukin-12 p35 subunit gene promoter sequence.
KW   interleukin-12; IL-12; p35 subunit; promoter; disease; treatment;
KW   cancer; viral infection; HIV; autoimmune; infectious; allergic; ds.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
misc_signal   1352
FT   /tag= a
FT   /note= "alternative transcription start site"
misc_signal   1594
FT   /tag= b
FT   /note= "alternative transcription start site"
misc_signal   1625
FT   /tag= c
FT   /note= "alternative transcription start site"
misc_signal   1626
FT   /tag= d
FT   /note= "alternative transcription start site"

PN   EP-818534-A1.
PD   14-JAN-1998.
PF   12-JUL-1996; 201977.
PR   12-JUL-1996; EP-201977.
PA   (JANC-) JANSSEN PHARM NV.
PI   De Chaffoy de Courcelles DRGG, Roevens PWM, Van Dijck PJG;
DR   WPI; 98-065306/07.
DR   DNA comprising first intron of IL-12 p35 subunit gene, useful for
PT   comprising first intron of IL-12 p40 subunit gene promoter - and
PT   screening for compounds to treat e.g. cancers, allergies,
PT   infections, etc
PS   Claim 2; Page 12-13; 23pp; English.
CC   The sequence is that of a human interleukin (IL)-12 p35 subunit gene
CC   promoter sequence. Cells containing a construct in which a reporter
CC   gene is placed under the control of the IL-12p35 and p40 subunit
CC   gene expression regulating sequences, can be used to screen for
CC   compounds which affect the production of IL-12 in cells, especially
CC   human cells. These compounds especially non-proteinaceous compounds
CC   different from lipopolysaccharide, can be used for the treatment of
CC   a disease involving a too high or too low IL-12 production, e.g.
CC   cancers, viral infections including HIV, immunodeficiencies,
CC   autoimmune diseases, infectious diseases or allergic diseases.
CC   The nucleic acid sequences can be used to make transgenic non-human
CC   mammalian animals which expresses a reporter gene under their control.
SQ   Sequence 1679 BP;   412 A;   408 C;   406 G;   452 T;

Qy   2 CTTAAGCTTGGAGTCAGTCACGTACGAGCAAGCT 31
        ||| |||||||||||| ||| ||| | | |||||
Db 1316 CTTTAGCTTGGGAGTGGTGGATGGGAAAAGCT 1345

Query Match   49.1%; Score 17.2; DB 1; Length 1679;
Best Local Similarity   73.3%; Pred. No. 57;
Matches   22; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

RESULT 15
ID   V74417
AC   V74417;
     V74417;
DT   16-MAR-1999 (first entry)
DE   Staphylococcus aureus contig SEQ ID #106.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
misc_feature   1321..1380
FT   /tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   3121..3180

RESULT 15
LOCUS   AQ690006/c
```
Class: BAC ends
High quality sequence stop: 413.
```

FEATURES   Location/Qualifiers
  source   1..545
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbx0096H07r"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world population especially those
                     inhabiting highly populated areas of the humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991). The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to facilitate positional cloning, physical mapping
                     and genome sequencing of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents.
                     The deep coverage allows the isolation a particular
                     sequence with a probability of 99.9 %. Two high density
                     filters, each containing 18,432 clones (doubly spotted),
                     represent the whole library for colony screening."

BASE COUNT   160 a   90 c   127 g   167 t   1 others
ORIGIN

Query Match   58.9%; Score 20.6; DB 79; Length 545;
Best Local Similarity 85.2%; Pred. No. 88;
Matches 23; Conservative   0; Mismatches 4; Indels 0; Gaps 0;

Qy   4 ATCTAGAATACTATGTTCTCACACTCTTA 30
        || ||| |||||||||||||||| | |||||
Db 397 ATGTAGGTACTATGTTCTATATTCTTA 371

```
High quality sequence stop: 228.
```

FEATURES   Location/Qualifiers
  source   1..364
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0081G11r"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world population especially those
                     inhabiting highly populated areas of the humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991). The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to facilitate positional cloning, physical mapping
                     and genome sequencing of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents.
                     The deep coverage allows the isolation a particular
                     sequence with a probability of 99.9 %. Two high density
                     filters, each containing 18,432 clones (doubly spotted),
                     represent the whole library for colony screening."

BASE COUNT   118 a   57 c   110 g   77 t   2 others
ORIGIN

Query Match   58.9%; Score 20.6; DB 81; Length 364;
Best Local Similarity 85.2%; Pred. No. 93;
Matches 23; Conservative   0; Mismatches 4; Indels 0; Gaps 0;

Qy   4 ATCTAGAATACTATGTTCTCACACTCTTA 30
        || ||| ||||||||||||||| | |||||
Db 206 ATGTAGGTACTATGTTCTATATTCTTA 180

LOCUS   AQ690006   364 bp   DNA   GSS   01-JUL-1999
DEFINITION nbxb0081G11r CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0081G11r, genomic survey sequence.
ACCESSION   AQ690006
NID   g5331174
VERSION   AQ690006.1 GI:5331174
KEYWORDS   GSS.
SOURCE   Oryza sativa.
  ORGANISM Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 364)
  AUTHORS   Yu,Y., Budiman,M.A., Boyd,C.A., Gaudette,B., Gayle,A.,
            Kingsbury,R., Phillips,K., Sasinowski,M., Dean,R.A. and Wing,R.A.
  TITLE   A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT   Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends

Search completed: December 29, 1999, 04:26:29
Job time: 2550 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: December 29, 1999, 04:26:29 ; Search time 727.51 Seconds
                                      (without alignments)
                                      164.915 Million cell updates/sec

Title:
Perfect score: 35
Sequence: 1 TCTTAAGCTTGGAGTCACGTACGTACGAGCAAGCTAGTT 35
Scoring table: IDENTITY_NUC
Searched: US-09-077-173A-4
Database :   4089388 seqs, 1713965092 residues
                 EST:*
Word size :   0

Number of hits that pass the threshold :   8178776

1: em_est1:*		50: gb_est31:*	
2: em_est2:*		51: gb_est32:*	
3: em_est3:*		52: em_est20:*	
4: em_est4:*		53: em_est21:*	
5: em_est5:*		54: em_est22:*	
6: em_est6:*		55: em_est23:*	
7: em_est7:*		56: em_est24:*	
8: em_est8:*		57: em_est25:*	
9: em_est9:*		58: em_est26:*	
10: em_est10:*		59: gb_est33:*	
11: em_est11:*		60: gb_est34:*	
12: em_est12:*		61: gb_est35:*	
13: em_est13:*		62: gb_est36:*	
14: em_est14:*		63: gb_est37:*	
15: em_est15:*		64: gb_est38:*	
16: em_est16:*		65: em_est27:*	
17: em_est17:*		66: em_est28:*	
18: em_est18:*		67: em_est29:*	
19: em_est19:*		68: em_est30:*	
20: gb_est1:*		69: gb_gss1:*	
21: gb_est2:*		70: gb_gss2:*	
22: gb_est3:*		71: gb_gss3:*	
23: gb_est4:*		72: gb_gss4:*	
24: gb_est5:*		73: em_gss1:*	
25: gb_est6:*		74: em_gss2:*	
26: gb_est7:*		75: em_gss3:*	
27: gb_est8:*		76: gb_gss5:*	
28: gb_est9:*		77: gb_gss6:*	
29: gb_est10:*		78: gb_gss7:*	
30: gb_est11:*		79: gb_gss8:*	
31: gb_est12:*		80: gb_gss8:*	
32: gb_est13:*		81: gb_gss9:*	
33: gb_est14:*		82: em_gss5:*	
34: gb_est15:*		83: em_gss6:*	
35: gb_est16:*		84: em_gss7:*	
36: gb_est17:*		85: em_gss8:*	
37: gb_est18:*		86: em_gss9:*	
38: gb_est19:*		87: em_gss10:*	
39: gb_est20:*		88: em_gss11:*	
40: gb_est21:*		89: gb_gss10:*	
41: gb_est22:*		90: gb_gss11:*	
42: gb_est23:*			
43: gb_est24:*			
44: gb_est25:*			
45: gb_est26:*			
46: gb_est27:*			
47: gb_est28:*			
48: gb_est29:*			
49: gb_est30:*			

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21.2	60.6	403	26	W34668	W34668 mc32e10.r1
2	21.2	60.6	507	29	AA049196	AA049196 mj49c10.r .
3	21.2	60.6	478	35	AA538294	AA538294 v198e05.r
4	21.2	60.6	446	43	AI158502	AI158502 ud25b03.r
5	20.4	58.3	260	39	AA895889	AA895889 vy35f03.r
6	20.2	57.7	424	21	T59835	T59835 yc13g07.r1 .
7	20.2	57.7	648	44	AI294097	AI294097 LP07395.5
8	20.2	57.7	453	81	AQ605448	AQ605448 HS_2119_B
9	20.2	57.7	410	23	H32055	H32055 EST106796_R
10	19.8	56.6	292	41	AI030851	AI030851 UI-R-C0-)
11	19.8	56.6	317	43	AI175808	AI175808 EST219379
12	19.8	56.6	493	43	AI229779	AI229779 EST226474
13	19.8	56.6	543	44	AI317823	AI317823 EST234494
14	19.8	56.6	443	46	AI028963	AI028963 UI-R-C0-1
15	19.8	56.6	564	46	AI409203	AI409203 EST37495
16	19.8	56.0	500	25	N57501	N57501 yy54e01.s1
17	19.6	56.0	366	46	AI438032	AI438032 sa35a07.y
18	19.6	56.0	351	46	AI440573	AI440573 sa84d11.y
19	19.6	56.0	345	46	AI440718	AI440718 sa84d11.y
20	19.6	56.0	419	46	AI441460	AI441460 sa86b12.y
21	19.6	56.0	398	46	AI443140	AI443140 sa84g12.y
22	19.6	56.0	616	61	AI855807	AI855807 sc29e11.y

```
RESULT    1
LOCUS       W34668       403 bp    mRNA
DEFINITION  mc32e10.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:350250 5', mRNA sequence.
ACCESSION   W34668
NID         g1316714
VERSION     W34668.1  GI:1316714
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 403)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:22050
            Seq primer: ETprimer
            High quality sequence stop: 400.
FEATURES             Location/Qualifiers
     source          1..403
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:350250"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
```

ALIGNMENTS

#					Accession	Name
23	19.6	56.0	380	62	AI930607	sb37b07.y
24	19.6	56.0	535	63	AI937978	sc06d11.y
25	19.4	55.4	813	41	AI068950	mgaee0004d
26	19.4	55.4	881	41	AI069341	mgaee0006c
27	19.4	55.4	143	72	AQ248417	AK016 Gen
28	19.2	54.9	670	81	AQ662079	HS_2093_B
29	19.2	54.9	581	81	AQ677595	HS_5516_A
30	19	54.3	285	61	AI845565	UI-M-AO1-
31	19	54.3	660	80	AQ398861	RPCI11-14
32	18.8	53.7	430	28	AA067065	mm31d10.r
33	18.8	53.7	354	30	AA268740	vb01b10.r
34	18.8	53.7	489	42	AI118320	ue35f06.x
35	18.8	53.7	542	49	AI652450	wb29c05.x
36	18.8	53.7	544	72	AQ263444	CITB1-El-
37	18.8	53.7	854	77	AQ288703	nbxb00330
38	18.8	53.7	726	78	AQ051139	nbxb0005a
39	18.8	53.7	540	80	AQ593998	HS_5303_A
40	18.8	53.7	554	90	AQ822473	HS_5547_B
41	18.8	53.7	399	27	AA054131	zf51b03.s
42	18.8	53.1	432	29	AA178090	mt06g07.r
43	18.6	53.1	418	46	AI442046	sa66h11.y
44	18.6	53.1	335	70	AQ012879	CIT-HSP-2
45	18.6	53.1	445	80	AQ381084	RPCI11-16

```
BASE COUNT      79 a    124 c     98 g    102 t
ORIGIN
        TGTTACCAAACTGAAGTGGGAGGCGGCCCGGCAnTTTTTTTTTTTTTTTTTTTT (3'),
        double-stranded cDNA was size selected, ligated to Eco RI
        adapters (Pharmacia), digested with Not I and cloned into
        the Not I and Eco RI sites of a modified pT7T3 vector
        (Pharmacia). Library went through one round of
        normalization to a Cot = 5. Library constructed by Bento
        Soares and M.Fatima Bonaldo. RNA was kindly provided by
        Dr. Minoru Ko (Wayne State University)."
```

```
RESULT    2
AA049196/c
LOCUS       AA049196    507 bp    mRNA    EST     30-DEC-1996
DEFINITION  mj49c10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:479442 5' similar to TR:E225817 E225817 HYPOTHETICAL
            64.0 KD PROTEIN. ;, mRNA sequence.
ACCESSION   AA049196
NID         g1755343
VERSION     AA049196.1  GI:1755343
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 507)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     On Dec 30, 1996 this sequence version replaced gi:1529025.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:290186
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1..507
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:479442"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc embryo"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCGGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
```

```
Query Match     79.8%;  Score 21.2;  DB 26;  Length 403;
Best Local Similarity   76.58%;  Pred. No. 12;
Matches   26;  Conservative   0;  Mismatches   8;  Indels   0;  Gaps   0;

Qy    1  TCTTAAGCTTGGAGTCACGTACGTACGAGCGAAGCTAGT 34
            ||| ||||| || ||| | || | ||||||||| |||
Db  330  TCTGAGGCTTGTAGGCCATGGAGGAGGAGGCCAAGCCCAGT 297
```

RESULT 3

Qy   1 TCTTAAGCTTGGAGTCACGTACGAGCAAGCTAGT 34
          ||| ||||||| ||||| || ||| | | |||||| |||
Db   416 TCTGAGGCTTGTGTAGGCATTGGAGGAGGCAAGCCAGT 383

Query Match   60.6%; Score 21.2; DB 29; Length 507;
Best Local Similarity   76.5%; Pred. No. 12;
Matches 26; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

LOCUS   AA538294/c
ACCESSION   AA538294
DEFINITION v198e05.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:920288 5', mRNA sequence.
LOCUS   AA538294   478 bp   mRNA   EST   29-JUL-1997
NID   g2284287
VERSION   AA538294.1 GI:2284287
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 478)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Oct 30, 1996 this sequence version replaced gi:1656984.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:532504
            Seq Primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 255.
FEATURES   Location/Qualifiers
     source   1..478
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:920288"
                     /clone_lib="Barstead mouse pooled organs MPLRB4"
                     /sex="mixed"
                     /tissue_type="pooled organs"
                     /dev_stage="7 day"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATTCGAAGTGGAGCGCGGCCGCCCTTTTTTTTTTTTTTTTTTT
                     3'], double-stranded cDNA was ligated to Eco RI adaptors
                     [GTTGGATTCGGTACC], digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector."
BASE COUNT   89 a   170 c   118 g   130 t
ORIGIN
      14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
      State Univ., from 2]; double-stranded cDNA was ligated to
      Eco RI adaptors (Pharmacia), digested with Not I and
      cloned into the Not I and Eco RI sites of the modified
      pT7T3 vector. Library went through one round of
      normalization, and was constructed by Bento Soares and
      M.Fatima Bonaldo. "

RESULT 4

Qy   1 TCTTAAGCTTGGAGTCACGTACCAGCAAGCTAGT 34
          ||| ||||||| ||| ||| | || | |||||| |||
Db   126 TCTGAGGCTTGTGTAGGCATTGGAGGAGGCAAGCCAGT 93

Query Match   60.6%; Score 21.2; DB 35; Length 478;
Best Local Similarity   76.5%; Pred. No. 12;
Matches 26; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

LOCUS   AI158502/c
ACCESSION   AI158502
DEFINITION ud25b03.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:1446893 5',
            mRNA sequence.
LOCUS   AI158502   446 bp   mRNA   EST   30-SEP-1998
NID   g3686971
VERSION   AI158502.1 GI:3686971
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 446)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2045702.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:920209
            Seq Primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 446.
FEATURES   Location/Qualifiers
     source   1..446
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="7; 15"
                     /clone="IMAGE:1446893"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATTCGAAGTGGAGCGCGGCCGCCGTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT   82 a   135 c   108 g   120 t   1 others
ORIGIN

BASE COUNT   94 a   141 c   117 g   126 t
ORIGIN   Library constructed by Bob Barstead. "

RESULT 5

Qy   1 TCTTAAGCTTTGGAGTCACGTACCGAGCAAGCCTAGT 34
         ||| |||||||||||||| || ||||||||| |||||| |||
Db 142 TCTGAGGCTTTGTTGTAGGCATGGAGGAGGCAAGCCAGT 109

Query Match   60.6%; Score 21.2; DB 43; Length 446;
Best Local Similarity 76.5%; Pred. No. 12;
Matches   26; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

LOCUS   AA895889   260 bp   mRNA   EST   06-APR-1998
DEFINITION vy35f03.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
            IMAGE:1297469 5' similar to gb:L07765 LIVER CARBOXYLESTERASE
            PRECURSOR (HUMAN); gb:M57960 Mouse carboxylesterase mRNA, complete
            cds (MOUSE);, mRNA sequence.
ACCESSION   AA895889
NID   g3032282
VERSION   AA895889.1 GI:3032282
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 260)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE   The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT   On Jan 17, 1998 this sequence version replaced gi:2044167.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source   1..260
                     /organism="Mus musculus"
                     /strain="C57BL/6 x CBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1297469"
                     /map="6"
                     /clone_lib="Stratagene mouse lung 937302"
                     /sex="female"
                     /tissue_type="lung"
                     /dev_stage="6-8 month old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     Oligo dT. 6-8 month old female lung and 1.5 year old male
                     lung were source of mRNA. Average insert size: 1.5 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'"
BASE COUNT   84 a   59 c   69 g   48 t
ORIGIN

Query Match   58.3%; Score 20.4; DB 39; Length 260;
Best Local Similarity 80.0%; Pred. No. 24;
Matches   24; Conservative   0; Mismatches   6; Indels   0; Gaps   0;

Qy   1 TCTTAAGCTTTGGAGTCACGTACCGAGCAAGCCTAGT 30
         ||| ||||||||||||||||||||| || |||||| |||||||
Db   43 TCTACAGATTGGACCCACTTACCAGCAAGC 72

RESULT 6

Qy   1 TCTTAAGCTTTGGAGTCACGTACCGAGCAAGC 30
         ||| |||||||||||||||| || |||||||| ||||||||
Db   43 TCTACAGATTGGACCCACTTACCAGCAAGC 72

Query Match   57.7%; Score 20.2; DB 21; Length 424;
Best Local Similarity 73.5%; Pred. No. 33;
Matches   25; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

LOCUS   T59835   424 bp   mRNA   EST   09-FEB-1995
DEFINITION yc13g07.r1 Stratagene lung (#937210) Homo sapiens cDNA clone
            IMAGE:80604 5', mRNA sequence.
ACCESSION   T59835
NID   g661672
VERSION   T59835.1 GI:661672
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 424)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE   Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT   Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 684
            High qality sequence stops: 220 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 684   Std Error: 0.00
            Seq primer: M13Rp1
            High quality sequence stop: 220.
FEATURES   Location/Qualifiers
     source   1..424
                     /organism="Homo sapiens"
                     /db_xref="GDB:484221"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:80604"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     Oligo dT. normal lung. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'"
                     3' -3' adaptor sequence: 5' GAATTCGGCACGAG
                     8 others
BASE COUNT   102 a   126 g   110 t
ORIGIN

Query Match   57.7%; Score 20.2; DB 21; Length 424;
Best Local Similarity 73.5%; Pred. No. 33;
Matches   25; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

Qy   2 CTTAAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35
         ||| |||||| |||||||||| ||||| ||||||| ||
Db 292 CTTTNGCTTGGGGTGGGTCAATTAGGAGAAGCTACTT 325

RESULT 7

LOCUS   AI294097   648 bp   mRNA   EST   01-DEC-1998
DEFINITION LP07395.5prime LP Drosophila melanogaster larval-early pupal pOT2
            Drosophila melanogaster cDNA clone LP07395 5prime, mRNA sequence.
ACCESSION   AI294097
NID   g3943504
VERSION   AI294097.1 GI:3943504
KEYWORDS   EST.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 648)
AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE   BDGP/HHMI Drosophila EST Project
JOURNAL   Unpublished (1997)
COMMENT   On Oct 17, 1997 this sequence version replaced gi:2518093.
            Contact: Harvey, D.
            G. M. Rubin:Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AL031583.
            Plate: 73 row: H column: 11
            High quality sequence stop: 546.
FEATURES   Location/Qualifiers
     source   1..648
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="22"
                     /clone_lib="LP07395"
                     /clone="LP07395 Drosophila melanogaster larval-early pupal
                     pOT2"
                     /sex="male and female"
                     /dev_stage="larvae-pupae"
                     /note="DH5-alpha"
                     /lab_host="DH5-alpha"
                     Site_2: XhoI; Site_1: EcoRI;
                     ligated into pOT2. Plasmid cDNA library."
BASE COUNT   180 a   150 c   151 g   167 t
ORIGIN

RESULT   8
AQ605448/c
LOCUS   AQ605448   453 bp   DNA   GSS   10-JUN-1999
DEFINITION HS_2119_B2_D07_MR CIT Approved Human Genomic Library D Homo
            sapiens genomic clone Plate-2119 Col-14 Row-H, genomic survey
            sequence.
ACCESSION   AQ605448
NID   g5065372
VERSION   AQ605448.1 GI:5065372
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 453)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,

Qy   1 TCTTAAGCTTGGAGTCACGTACGACCAAGCTAG 33
        ||||||| |||||||||||||||||||||||||
Db 501 TCTTGATGTTCGTGTCTCACCAATGAGCAAGATAG 533

Query Match   57.7%;   Score 20.2; DB 44; Length 648;
Best Local Similarity   75.8%; Pred. No. 37;
Matches   25; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

JOURNAL   Shaker,R., Schmidt,S., Tralcoff,R. and Hood,L.E.
TITLE   Construction of a Characterized Clone Resource for Genomic
            Sequencing
COMMENT   Unpublished '(1998)
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2119 row: H column: 14
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 453.
FEATURES   Location/Qualifiers
     source   1..453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate-2119 Col-14 Row=H"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT   111 a   95 c   89 g   155 t   3 others
ORIGIN

Qy   1 TCTTAAGCTTGGAGTCACGTACGACCAAGCTAG 33
        ||||||||||| || ||||| || ||||| |||
Db   56 TCTTAAGCTGAGAGTCATGTTCCAGAATGCCAG 24

Query Match   57.7%;   Score 20.2; DB 81; Length 453;
Best Local Similarity   75.8%; Pred. No. 34;
Matches   25; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

RESULT   9
H32055
LOCUS   H32055   410 bp   mRNA   EST   02-APR-1998
DEFINITION EST106796 Rat PC-12 cells, untreated Rattus sp. cDNA clone RPCBG19
            5' end, mRNA sequence.
ACCESSION   H32055
NID   g977472
VERSION   H32055.1 GI:977472
KEYWORDS   EST.
SOURCE   Rattus sp.
ORGANISM   Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 410)
AUTHORS   Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D.,
            Kelavage,A.R., Fraser,C.M. and Venter,J.C.
TITLE   Comparative expressed-sequence-tag analysis of differential gene
            expression profiles in PC-12 cells before and after nerve growth
            factor treatment
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
MEDLINE   95396786
COMMENT   Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            For clone availability please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.

RESULT 10

```
FEATURES
      source   Location/Qualifiers
               1..410
               /organism="Rattus sp."
               /db_xref="ATCC (inhost):2001506"
               /clone="RPCBG19"
               /db_xref="taxon:10118"
               /clone_lib="Rat Pc-12 cells, untreated"
               /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
               XhoI; poly(A)+ RNA was purified from untreated PcI2 cells
               cultured for 9 days; cDNA was constructed using an
               oligo-dT primer and directionally cloned using the Lambda
               ZAP II Vector Kit by Stratagene"

BASE COUNT      83 a     85 c    118 g    123 t     1 others
ORIGIN

Qy       5  AAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35
            |||||||||||||| || ||||||| ||||||
Db     244  AAGCTGGGAGTAACTTACTCGGAAGCTAATT 274
```

RESULT 10

```
RESULT 10
AI030851/c
LOCUS       AI030851         292 bp    mRNA          EST    11-FEB-1999
DEFINITION  UI-R-CO-jp-b-03-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone
            UI-R-CO-jp-b-03-0-UI 3', mRNA sequence.
ACCESSION   AI030851
NID         g3248677
VERSION     AI030851.1    GI:3248677
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2044130.

            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            On the sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics The following repetitive elements were found in
            this cDNA sequence: 1-44, >POLY_A#Simple_repeat
            Seq primer: M13 Forward.

FEATURES
      source   Location/Qualifiers
               1..292
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /map="4"
               /clone="UI-R-CO-jp-b-03-0-UI"
               /clone_lib="UI-R-CO"
               /dev_stage="adult"
               /lab_host="DH10B (Life Technologies)"
               /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
               polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AI
               library is a subtracted library derived from the UI-R-AI
```

and UI-R-El libraries. The UI-R-Al library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, and muscle. The UI-R-El
library consisted of a mixture of individually tagged
normalized libraries constructed from 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dT track which
allows identification of the library of origin of a clone
within the mixture. The subtracted library (UI-R-CO) was
constructed as follows: PCR amplified cDNA inserts from a
pool of UI-R-Al and UI-R-El clones from which 3' ESTs had
been derived was used as a driver in a hybridization with
the pooled UI-R-Al and UI-R-El library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-CO
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."

RESULT 11

```
BASE COUNT      92 a     70 c     43 g     87 t
ORIGIN

Qy       5  AAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35
            |||||||||||||| || ||||||| || || |
Db     214  AAGCTGGGAGTAACTTACTCGGAAGCTAATT 184
```

RESULT 11

```
RESULT 11
AI175808/c
LOCUS       AI175808         317 bp    mRNA          EST    20-JAN-1999
DEFINITION  EST219379 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVBF72 3' end, mRNA sequence.
ACCESSION   AI175808
NID         g3726446
VERSION     AI175808.1    GI:3726446
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Lee,N.H., Glodek,A.R., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2045206.
            Other_ESTs: TC49823
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.

FEATURES
      source   Location/Qualifiers
               1..317
               /organism="Rattus sp."
               /db_xref="ATCC (inhost):2032482"
               /db_xref="taxon:10118"
               /clone="ROVBF72"
               /clone_lib="Normalized rat ovary, Bento Soares"
               /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
               Site_2: NotI"
```

RESULT 12

```
Query Match        Best Local Similarity  56.6%;   Score 19.8;   DB 43;   Length 317;
Matches  24;   Conservative  0;   Mismatches  7;   Indels  0;   Gaps  0;
                                     Pred. No. 47;

Qy    5 AAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35
        |||||  ||||| ||| |  || |||||||| ||
Db  195 AAGCTGGGAGTAACTTACTGGGAAGCTAATT 165
```

```
LOCUS        AI229779/c
DEFINITION   AI229779  493 bp  mRNA   EST   20-JAN-1999
             REMCL85 3' end, mRNA sequence.
ACCESSION    AI229779
NID          AI229779.1  GI:3813666
VERSION      g3813666
KEYWORDS     EST.
SOURCE       Rattus sp.
ORGANISM     Rattus sp.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE    1 (bases 1 to 493)
AUTHORS      Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
             Kerlavage,A.R. and Adams,M.D.
TITLE        Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
             Gene Index
JOURNAL      Unpublished (1998)
COMMENT      On Aug 21, 1998 this sequence version replaced.

             Contact: Lee, NH
             ATCC
             The Institute for Genomic Research
             9712, Medical Center Drive, Rockville, MD 20850, USA
             Tel: (301)-838-3529
             Fax: (301)-838-0208
             Email: nhlee@tigr.org
             Seq primer: M13-21.
FEATURES     Location/Qualifiers
   source    1..493
             /organism="Rattus sp."
             /db_xref="ATCC (inhost):2040031"
             /db_xref="taxon:10118"
             /clone="REMCL85"
             /clone_lib="Normalized rat embryo, Bento Soares"
             /dev_stage="embryo 8,12,18 dpc"
             /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT    105 a    86 c    52 g    74 t
ORIGIN
```

RESULT 13

```
Query Match        Best Local Similarity  56.6%;   Score 19.8;   DB 43;   Length 543;
Matches  24;   Conservative  0;   Mismatches  7;   Indels  0;   Gaps  0;
                                     Pred. No. 53;

Qy    5 AAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35
        |||||  ||||| ||| |  || |||||||| ||
Db  194 AAGCTGGGAGTAACTTACTGGGAAGCTAATT 164
```

```
LOCUS        AI317823/c
DEFINITION   AI317823  543 bp  mRNA   EST   17-DEC-1998
             EST234494 PC12 cells, untreated, pT7T3Pac, TIGR Rattus sp. cDNA
             clone RPPAA31 3' end, mRNA sequence.
ACCESSION    AI317823
NID          g4033090
VERSION      AI317823.1  GI:4033090
KEYWORDS     EST.
SOURCE       Rattus sp.
ORGANISM     Rattus sp.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE    1 (bases 1 to 543)
AUTHORS      Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
             Kerlavage,A.R. and Adams,M.D.
TITLE        Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
             Gene Index
JOURNAL      Unpublished (1998)
COMMENT      On Jan 19, 1998 this sequence version replaced gi:2151434.

             Contact: Lee, NH
             ATCC
             The Institute for Genomic Research
             9712, Medical Center Drive, Rockville, MD 20850, USA
             Tel: (301)-838-3529
             Fax: (301)-838-0208
             Email: nhlee@tigr.org
             Seq primer: M13-21.
FEATURES     Location/Qualifiers
   source    1..543
             /organism="Rattus sp."
             /db_xref="taxon:10118"
             /clone="RPPAA31"
             /clone_lib="PC12 cells, untreated, pT7T3Pac"
             /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT    157 a    116 c    113 g    107 t
ORIGIN
```

RESULT 14

```
Query Match        Best Local Similarity  56.6%;   Score 19.8;   DB 44;   Length 543;
Matches  24;   Conservative  0;   Mismatches  7;   Indels  0;   Gaps  0;
                                     Pred. No. 54;

Qy    5 AAGCTTGGAGTCACGTACGAGCAAGCTAGTT 35
        |||||  ||||| ||| |  || |||||||| ||
Db  190 AAGCTGGGAGTAACTTACTGGGAAGCTAATT 160
```

```
LOCUS        AI028963/c
DEFINITION   AI028963  443 bp  mRNA   EST   04-JUL-1999
             UI-R-CO-1o-f-11-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone
             UI-R-CO-1o-f-11-0-UI 3', mRNA sequence.
ACCESSION    AI028963
NID          g4295273
VERSION      AI028963.1  GI:4295273
KEYWORDS     EST.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE    1 (bases 1 to 443)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
             discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
COMMENT      On Jun 22, 1998 this sequence version replaced gi:3246789.

             Contact: Soares, MB
             Program for Rat Gene Discovery and Mapping
             University of Iowa
             451 Eckstein Medical Research Building Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: msoares@blue.weeg.uiowa.edu
             The sequence tag present in the cDNA between the NotI site and the
             oligo-dT track served to identify it as a clone from the normalized
             adult placenta library, cDNA Library Preparation: M. Fatima
             Bonaldo, Ph.D. Clone distribution: clones will be available through
BASE COUNT    165 a    154 c    107 g    117 t
ORIGIN
```

FEATURES
source

Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
ID=174389 The following repetitive elements were found in this
cDNA sequence: 1-42, >POLY_A#Simple_repeat
Seq primer: M13 Forward
POLYA=No.

1. .443
  Location/Qualifiers
  /organism="Rattus norvegicus"
  /strain="Sprague-Dawley"
  /db_xref="taxon:10116"
  /clone="UI-R-C0-lo-f-11-0-UI"
  /clone_lib="UI-R-C0"
  /dev_stage="adult"
  /lab_host="DH10B (Life Technologies)"
  /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
  polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
  library is a subtracted library derived from the UI-R-A1
  and UI-R-E1 libraries. The UI-R-A1 library consisted of a
  mixture of individually tagged normalized libraries
  constructed from rat placenta, adult lung, brain, liver,
  kidney, heart, spleen, ovary, and muscle. The UI-R-E1
  library consisted of a mixture of individually tagged
  normalized libraries constructed from 8, 12 and 18-day
  embryo. The tag is a string of 3-5 nucleotides present
  between the Not I site and the oligo-dT track which
  allows identification of the library of origin of a clone
  within the mixture. The subtracted library (UI-R-C0) was
  constructed as follows: PCR amplified cDNA inserts from a
  pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
  been derived was used as a driver in a hybridization with
  the pooled UI-R-A1 and UI-R-E1 library in the form of
  single-stranded circles. The remaining single-stranded
  circles (subtracted library) was purified by
  hydroxyapatite column chromatography, converted to
  double-stranded circles and electroporated into DH10B
  bacteria (Life Technologies) to generate the UI-R-C0
  library. This procedure has been previously described
  (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
  1996)."

BASE COUNT   128 a   119 c   84 g   112 t
ORIGIN

RESULT 15
LOCUS   AI409203/c
AI409203
DEFINITION   EST237495 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
             RKIDN27 3' end, mRNA sequence.
ACCESSION   AI409203
NID   g4252707
VERSION   AI409203.1 GI:4252707
KEYWORDS   EST.
SOURCE   Rattus sp.
  ORGANISM   Rattus sp.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 564)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
             Kerlavage,A.R. and Adams,M.D.
  TITLE   Rat Gene Project: Generation of a Rat EST (REST) Catalog & Rat
             Gene Index
  JOURNAL   Unpublished (1998)
COMMENT   On Mar 10, 1998 this sequence version replaced gi:2949295.

Query Match
Best Local Similarity   56.6%;
Matches   24; Conservative   0;   Mismatches   7;   Indels   0;   Gaps   0;

Qy   5 AAGCTTGGAGTCACGTACGAGCAAAGCTAGTT 35
           |||||| ||||| || ||| | ||||||||| ||
Db   212 AAGCTGGGAGTAACTTACTGGGAAGCTAATT 182

Score 19.8;   DB 46;   Length 443;
Pred. No. 51;

Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source
  1. .564
  Location/Qualifiers
  /organism="Rattus sp."
  /db_xref="taxon:10118"
  /clone="RKIDN27"
  /clone_lib="Normalized rat kidney, Bento Soares"
  /note="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
  Site_2: NotI"

BASE COUNT   171 a   156 c   112 g   125 t
ORIGIN

Query Match
Best Local Similarity   56.6%;
Matches   24; Conservative   0;   Mismatches   7;   Indels   0;   Gaps   0;

Qy   5 AAGCTTGGAGTCACGTACGAGCAAAGCTAGTT 35
           |||||| ||||| || ||| | ||||||||| ||
Db   195 AAGCTGGGAGTAACTTACTGGGAAGCTAATT 165

Score 19.8;   DB 46;   Length 564;
Pred. No. 54;

Search completed: December 29, 1999, 04:26:33
Job time: 2554 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: December 29, 1999, 07:18:38 ; Search time 601.47 Seconds
(without alignments)
248.818 Million cell updates/sec

Title: US-09-077-173a-3
Perfect score: 35
Sequence: 1 CAGATCTAGAATACTATGTTCTPACACTCTTACGTGC 35
Scoring table: IDENTITY_NUC
Searched: 780561 seqs, 2137953050 residues
Database :   GenEmbl:*
Word size :   0

Number of hits that pass the threshold : 1561122

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_htg:*
19: em_htg1:*
20: em_hum2:*
21: em_in:*
22: em_om:*
23: em_or:*
24: em_ov:*
25: em_pat:*
26: em_ph:*
27: em_pl:*
28: em_pr:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*
43: gb_htg3:*
44: gb_htg4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	35	100.0	35	5	A62987	A62987 Sequence 3
2	22	62.9	226345	41	AC005406	AC005406 , complet
3	21.6	61.7	70952	7	AB024028	AB024028 Arabidops
4	21.2	60.6	282610	1	RPXX01	AJ235270 Ricketts1
5	21.2	60.6	78921	7	AB028621	AB028621 Arabidops
6	21.2	60.6	212363	34	AC009499	AC009499 Homo sapi
7	20.8	59.4	94817	10	HS981L23	AL031686 Human DNA
8	20.8	59.4	1790	10	AF049125	AF049125 Mus muscu
9	20.6	58.9	235	13	G04095	G04095 human STS W
10	20.2	57.7	5701	7	SPCC1902	AL049521 S.pombe C
11	20.2	57.7	203006	10	HS1145L23	AL096699 Human DNA
12	20.2	57.7	147102	11	AC002476	AC002476 Human PAC
13	20.2	57.7	112359	11	HS060822	U60822 Human dystr
14	20.2	57.7	439	11	G21425	G21425 human STS W
15	20.2	57.7	163384	34	AC006962	AC006962 Homo sapi
16	20.2	57.7	112734	35	CEH16D19	AL031624 Caenorhab
17	20.2	57.7	95272	36	AC005298	AC005298 Drosophil
18	20	57.1	10167	2	MPAE000035	AE000035 Mycoplasm
19	20	57.1	184886	11	AC005358	AC005358 Homo sapi
20	20	57.1	21352	43	AC011659	AC011659 Homo sapi
21	20	57.1	72872	8	F1104	AF096370 Arabidops
22	19.8	56.6	1825	12	AB028143	AB028143 Arabidops
23	19.8	56.6	33071	42	AC011030	AC011030 Mus muscu
24	19.8	56.6	217829	43	AC011361	AC011361 Homo sapi
25	19.6	56.0	74903	8	ATAC005312	AC005312 Arabidops
26	19.6	56.0	1113457	11	AC003686	AC003686 Homo sapi
27	19.6	56.0	148676	33	AF029367	AF029367 Homo sapi
28	19.6	56.0	179285	33	CNS01DS6	AL121656 Homo sapi
29	19.6	56.0	220000	34	AC007126	AC007126 Homo sapi
30	19.6	56.0	133318	34	AC007770	AC007770 Homo sapi
31	19.6	56.0	124646	34	AC007822	AC007822 Drosophil
32	19.4	55.4	12984	2	AE001122	AE001122 Borrelia
33	19.4	55.4	78600	7	SCDNCH2	X79489 S.cerevisia
34	19.4	55.4	1425	7	SCROX3	X58300 S.cerevisia
35	19.4	55.4	2447	7	SCVBL092W	Z35853 S.cerevisia
36	19.4	55.4	1880	7	SCVBL093C	Z35854 S.cerevisia
37	19.4	55.4	167075	10	HS742J24	AL035448 Human DNA
38	19.4	55.4	143453	10	HS2561G12	AL049797 Human DNA
39	19.4	55.4	1203	10	HUMDYSKW	M86524 Human dystr
40	19.4	55.4	38770	10	HUMDYSTROP	M65036 Human dystr
41	19.4	55.4	7889	16	AF020905	AF020905 Common ch
42	19.4	55.4	241519	33	HS1012F16	AL080274 Homo sapi
43	19.4	55.4	91745	33	HSBG13C1	AL096802 Homo sapi
44	19.4	55.4	87319	34	AC004378	AC004378 Drosophil
45	19.4	55.4	37494	36	CELC04E12	AF078785 Caenorhab

ALIGNMENTS

RESULT 1
A62987

LOCUS   A62987   35 bp   DNA   PAT   12-MAR-1998
DEFINITION   Sequence 3 from Patent WO9719170.
ACCESSION   A62987
NID   g3716859
VERSION   A62987.1   GI:3716859
KEYWORDS   .
SOURCE   unidentified.
  ORGANISM   unidentified
             unclassified.
REFERENCE   1 (bases 1 to 35)
  AUTHORS   Communi,D., Pirotton,S., Parmentier,M. and Boeynaems,J.
  TITLE   RECEPTOR AND NUCLEIC ACID MOLECULE ENCODING SAID RECEPTOR
  JOURNAL   Patent: WO 9719170-A 3 29-MAY-1997;
             EUROSCREEN S A (BE)
FEATURES   Location/Qualifiers
     source   1..35
                     /organism="unidentified"

BASE COUNT   9 a   9 c   5 g   12 t
ORIGIN   /db_xref="taxon:32644"

Qy   1 CAGATCTAGATACTATGTTCTACACACTTACGTGC 35
            |||||||||||||||||||||||||||||||||||
Db   1 CAGATCTAGATACTATGTTCTACACACTTACGTGC 35

Query Match   100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 8.5e-05;
Matches   35; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT   2

LOCUS   AC005406   226345 bp   DNA   PRI   24-APR-1999
DEFINITION Homo sapiens 22q11 BAC Clone b128f11 In CES Region
            , complete sequence.
ACCESSION   AC005406
NID   g4580403
VERSION   AC005406.2 GI:4580403
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-AUG-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-SEP-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-SEP-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-OCT-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-OCT-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   8 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-NOV-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   9 (bases 1 to 226345)

REFERENCE   10 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-NOV-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   11 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-NOV-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   12 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-JAN-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   13 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-NOV-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   14 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-JAN-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   15 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-APR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   16 (bases 1 to 226345)
  AUTHORS   Fu,Y., Pan,H., McDermid,H. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-APR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT   On Apr 9, 1999 this sequence version replaced gi:3928116.
FEATURES   Location/Qualifiers
     source   1..226345
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   70721 a   47619 c   43447 g   64558 t
ORIGIN

Query Match   62.9%; Score 22; DB 41; Length 226345;
Best Local Similarity   83.3%; Pred. No. 6.3;
Matches   25; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   1 CAGATCTAGATACTATGTTCTACACACTTA 30
            |||||||||| |||| |||| |||||| ||
Db   174495 CAGATCTAGATGATTCTATCATCAACACACTGTTA 174524

RESULT   3

LOCUS   AB024028/c   70952 bp   DNA   PLN   03-MAR-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone: K1G2,
            complete sequence.
ACCESSION   AB024028
NID   g4519187
VERSION   AB024028.1 GI:4519187

```
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:KIG2.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1 (sites)
  AUTHORS   Nakamura,Y.
  TITLE     Structural Analysis of Arabidopsis thaliana Chromosome 3. I
  JOURNAL   Unpublished (1998)
REFERENCE   2 (bases 1 to 70952)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
            Fax:+81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..70952
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="KIG2"
                     /clone_lib="Mitsui P1"
BASE COUNT    23075 a  12412 c  12631 g  22834 t
ORIGIN
```

```
Query Match             61.7%;  Score 21.6;  DB 7;  Length 70952;
  Best Local Similarity   85.7%;  Pred. No. 11;
  Matches   24;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

QY          1 CAGATCTAGAATACTATGTTCTACACTCT 28
              ||  ||||||||||||||||||||| ||| |
Db      53707 CACGTCTGAATACTATGTTCTAAAACTAT 53680
```

```
RESULT  4
RPXX01/c
LOCUS       RPXX01     282610 bp    DNA             BCT       11-NOV-1998
DEFINITION  Rickettsia prowazekii strain Madrid E, complete genome; segment
            1/4.
ACCESSION   AJ235270 AJ235269
NID         g3860572
VERSION     AJ235270.1  GI:3860572
KEYWORDS    complete genome.
SOURCE      Rickettsia prowazekii
  ORGANISM  Rickettsia prowazekii
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE   1 (bases 1 to 282610)
  AUTHORS   Andersson,S.G., Zomorodipour,A., Andersson,J.O.,
            Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K.,
            Eriksson,A.S., Winkler,H.H. and Kurland,C.G.
  TITLE     The genome sequence of Rickettsia prowazekii and the origin of
            mitochondria
  JOURNAL   Nature 396 (6707), 133-140 (1998)
  MEDLINE   99039499
REFERENCE   2 (bases 1 to 282610)
  AUTHORS   Andersson,S.G.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1998) S.G.E. Andersson,
            Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University
            of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
FEATURES             Location/Qualifiers
     source          1..282610
                     /organism="Rickettsia prowazekii"
                     /strain="Madrid E"
                     /db_xref="taxon:782"
```

```
     gene            1..822
                     /gene="RP001"
     CDS             1..822
                     /gene="RP001"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA14474.1"
                     /db_xref="PID:e1343317"
                     /db_xref="GI:3860573"

     gene            906..1298
                     /gene="RP002"
     CDS             906..1298
                     /gene="RP002"
                     /codon_start=1
                     /transl_table=11
                     /product="THIOREDOXIN (trxA)"
                     /protein_id="CAA14475.1"
                     /db_xref="PID:e1343318"
                     /db_xref="GI:3860574"
                     /translation="MSCVNEITTLLEFDSNDINTTQRINMVNNVNTDSSFKNEVLESDL
                     PVMVDFWAEWCGPCKMLIPIIDEISKELQDKVVLKMIDENPKTPSEYGIRSIPTIM
                     LFKNGEQKDTKIGLQQKNSLLDWINKSI"

     gene            1309..2067
                     /gene="RP003"
     CDS             1309..2067
                     /gene="RP003"
                     /codon_start=1
                     /transl_table=11
                     /product="O-ANTIGEN EXPORT SYSTEM ATP-BINDING PROTEIN RFBE
                     (rfbE)"
                     /protein_id="CAA14476.1"
                     /db_xref="PID:e1343319"
                     /db_xref="GI:3860575"
                     /translation="MLLNHSKVFEITDGYVEGTIDVHKRAQGLKHFFLKKGVSLSPTI
                     PILNNINPSCYEGEKIAFIGSNGSGKSSLLKLIAGIYPLKSGIVKVHGDIAAIIDMGV
                     GFEQDTGRENIKMLMLYNMLDKYSKKIEKEIIDFSELGSKIDLPIKIYSSGMLSRL
                     AFSVSIFQNPQILLLDEYFAAGDSYFIEKSLNLMKNKFKNTPISIIVSHQEEIKDNC
                     DRCILLKDGHIDDGTPSEIFKIYKQQSNKEIHK"

     gene            2064..2840
                     /gene="RP004"
     CDS             2064..2840
                     /gene="RP004"
                     /codon_start=1
                     /transl_table=11
                     /product="O-ANTIGEN EXPORT SYSTEM PERMEASE PROTEIN RFBA
                     (rfbA)"
                     /protein_id="CAA14477.1"
                     /db_xref="PID:e1343320"
                     /db_xref="GI:3860576"
                     /translation="MIKYFFSKKYWRAIALLVKASIVRQNKDSFLGSLWSLLQPFIHI
                     IVISYFGFLLRQPREFMWYNLVGGIPLMSFIVSSLTICASSLVTRDQIIKKVRISKT
                     FFPIADSLGQLYTLICSFSAMYFAFILLFPEKFSWQIIFMPILLLPLVICVISGSIAV
                     INGSLPDMMLIIKSVIVAFITVVISYLIYRQFSKRVIYVL"

     gene            2837..3724
                     /gene="RP005"
     CDS             2837..3724
                     /gene="RP005"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA14478.1"
                     /db_xref="PID:e1343321"
```

```
                     /db_xref="PID:e1342325"
                     /db_xref="PID:g3860581"
                     /db_xref="GI:3860581"
                     /translation="MVSSNFYKNLGPRKLTAIVDFLHDIIEPTKIYEDIIDYIKILQ
                     EASPNDISFLSNPKYSEFLKTTKAAACIVPKNFTEEVNONTVLHAAENSYFAVSKLID
                     FFYAPIKISYTKIMKSAIIADSATIGKNCYIGHNVIEDDVIIGDNSIIDAGTFIGRG
                     VNIGKNARIEQHVSINIYALIIGDDVVILVAKGIGQDGFSTEKGVHHKIFHIGIVKIG
                     NNVEIGSNTIIDRGALODDIIEDLCRIDNLVQIGHGVKIGKGSIIVAQAGIAGSSAIG
                     KYCALGGQVGIAGHLNIGDGTQVAAQGGVAQNIEECKIVGGSPAVPIMDWHRQSTIMK
                     QLVKTSNSKLKK"

     tRNA            complement(9267..9339)
                     /gene="tRNA Phe"
                     /product="tRNA Phe (GAA)"
     gene            complement(9267..9339)
                     /gene="tRNA Phe"
     CDS             9661..10686
                     /gene="RP011"
                     /codon_start=1
                     /transl_table=11
                     /product="NIFR3-LIKE PROTEIN (nifR3)"
                     /protein_id="CAA14483.1"
                     /db_xref="PID:e1342326"
                     /db_xref="PID:g3860582"
                     /db_xref="GI:3860582"

     gene            3757..5877
                     /gene="RP006"
     CDS             3757..5877
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA14479.1"
                     /db_xref="PID:e1342322"
                     /db_xref="PID:g3860578"
                     /db_xref="GI:3860578"
                     /translation="MVAATGHRICNDCSKACIYQKQDEVNIPLIESNILEETLKLPYG
                     LEIYILLTRWNPLNIYAPLPKPEPTNYNILVTGLGPAGFSLSYLLRSGHNVIAIDGLK
                     IIPLSFNIHKPVFWFHEYKNLLSERIPRKGFGGVAEYGITIRWDKNNDLIRLILERNN
                     NFKYDGVALGENITKEQALDLSPDHVAPCIGAGKPKILNIENFEAKGVRTASDFLMT
                     QNYSKDIRKLIEPVALSPADIDIEKGLISFIVEVGKSTSLCKTLSENEKVVLMGPTG
                     SPLEIPONKKIIIIDSKYRNVGLLKILKENKNKVIFATYPDIKKHKLTSVDIVINIS
                     PEIAEEELQELKIFGENTELIINVNSLMQCMMKGICGQCIQKVKGKQKIFACSEQNQN
                     VEIIDFKSLKTRLRQNSLQEKMSN"

     gene            complement(6227..7021)
                     /gene="RP007"
     CDS             complement(6227..7021)
                     /codon_start=1
                     /transl_table=11
                     /product="ACYL-[ACYL-CARRIER-PROTEIN]-UDP-N-
                     ACETYLGLUCOSAMINE O-ACYLTRANSFERASE (lpxA)."
                     /protein_id="CAA14480.1"
                     /db_xref="PID:e1342323"
                     /db_xref="PID:g3860579"
                     /db_xref="GI:3860579"
                     /translation="MSNSNIHTTAIIAEAGAKFGKNVKVGPVICIIGPEVVLHDNVELKS
                     HVVIDGTIEIGENTVIVPPASICQPQPQILKYANERSSTIIGSNNTIREYVTVQAGSKS
                     GMITRVGNNNLPMVGVHIGHDCKIGNNLVFANYVSLAGHIKVGDYAIIGGLSAVHQY
                     TRIGEYSMIGGLSPVGADVIPFGLVSSKRAVLEGINLIGMRKGFDKADSLTALNAVE
                     EIFLGEGNPVDRIKQVAEKYKNNSIVTQIDIFLNQDSSRAFCHFKK"

     gene            complement(7031..7468)
                     /gene="RP008"
     CDS             complement(7031..7468)
                     /codon_start=1
                     /transl_table=11
                     /product="(3R)-HYDROXYMYRISTOYL-[ACYL CARRIER PROTEIN]
                     DEHYDRATASE (fabZ)"
                     /protein_id="CAA14481.1"
                     /db_xref="PID:e1342324"
                     /db_xref="PID:g3860580"
                     /db_xref="GI:3860580"
                     /translation="MIAITEIMDLIPHRYPELLVDRVLKIDPNKSIIGIKNVTVNEP
                     QFTGHFPARPVMPGVLMVESMAQLAALIVAKSLDSTKNKEVFLMSIENTKFRIVQPG
                     DTMHIHSVIDQQRANVWKFSSKVMVECEIAAESKFTAMIKDKM"

     gene            complement(7659..8699)
                     /gene="RP009"
     CDS             complement(7659..8699)
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-3-O-[3-HYDROXYMYRISTOYL] GLUCOSAMINE
                     N-ACYLTRANSFERASE (lpxD)"
                     /protein_id="CAA14482.1"
```

```
RESULT 5
LOCUS       AB028621/c
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MJJ8,
            complete sequence.
ACCESSION   AB028621
NID         g5041974
VERSION     AB028621.1  GI:5041974
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MJJ8.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Nakamura,Y.
  TITLE     Structural Analysis of Arabidopsis thaliana Chromosome 3. II
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 78921)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1999) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3934,
            Fax:+81-438-52-3994)
FEATURES             Location/Qualifiers
     source          1..78921
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="MJJ8"
                     /clone_lib="Mitsui P1"
BASE COUNT    26493 a  14014 c  13288 g  25126 t
ORIGIN
```

```
Qy     2 AGATCTAGATACTATGTTCTACACTCTTACGTGC 35
           |||| |||| ||| ||||| || |||||| |||||
Db 32913 AAATCTTAATAATATCTTCACTCTTTTAAGTGC 32880

Query Match             60.6%;  Score 21.2;  DB 1;  Length 282610;
Best Local Similarity   76.5%;  Pred. No. 13;
Matches   26;  Conservative   0;  Mismatches   8;  Indels   0;  Gaps   0;
```

Qy   1 CAGATCTAGATACTAGTTCTACACTCTTACGTG 34

RESULT   6

Query Match   60.6%;   Score 21.2;   DB 7;   Length 78921;
Best Local Similarity   76.5%;   Pred. No. 16;
Matches   26;   Conservative   0;   Mismatches   8;   Indels   0;   Gaps   0;

Db   57900 CAAATCTAGATTATATGTTCTCTACTAATAATGTG 57867

Qy   1 CAGATCTAGATACTAGTTCTACACTCTTACGTG 34
        ||| |||||||| ||||||| |||| ||| |||
Db   57900 CAAATCTAGATTATATGTTCTCTACTAATAATGTG 57867

ACCESSION   AC009499
LOCUS   AC009499   212363 bp   DNA   HTG   24-AUG-1999
DEFINITION Homo sapiens clone NH0510D10, WORKING DRAFT SEQUENCE, 16 unordered
            pieces.
ACCESSION   AC009499
VERSION   AC009499.1 GI:5762568
NID   g5762568
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 212363)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 212363)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   NOTE: This is a 'working draft' sequence. It currently
            consists of 16 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            1   1987: contig of 1987 bp in length
                  1988   2005: gap of unknown length
                  2006   4154: contig of 2149 bp in length
                  4155   4172: gap of unknown length
                  4173   6140: contig of 1968 bp in length
                  6141   6158: gap of unknown length
                  6159   8980: contig of 2822 bp in length
                  8981   8998: gap of unknown length
                  8999   11264: contig of 2266 bp in length
                 11265   11282: gap of unknown length
                 11283   14408: contig of 3126 bp in length
                 14409   14426: gap of unknown length
                 14427   17873: contig of 3447 bp in length
                 17874   17891: gap of unknown length
                 17892   22734: contig of 4843 bp in length
                 22735   22752: gap of unknown length
                 22753   28660: contig of 5908 bp in length
                 28661   28678: gap of unknown length
                 28679   34714: contig of 6036 bp in length
                 34715   34732: gap of unknown length
                 34733   43061: contig of 8329 bp in length
                 43062   43079: gap of unknown length
                 43080   51186: contig of 8107 bp in length
                 51187   51204: gap of unknown length
                 51205   73417: contig of 22213 bp in length
                 73418   73435: gap of unknown length
                 73436   101632: contig of 28197 bp in length
                101633   101650: gap of unknown length
                101651   152551: contig of 50901 bp in length
                152552   152569: gap of unknown length
                152570   212363: contig of 59794 bp in length.
FEATURES   Location/Qualifiers

BASE COUNT   67114 a   41483 c   40358 g   63138 t   270 others
ORIGIN

source   1..212363
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="NH0510D10"

Qy   1 CAGATCTAGATACTAGTTCTACACTCTTACGTG 34

RESULT   7

Query Match   60.6%;   Score 21.2;   DB 34;   Length 212363;
Best Local Similarity   76.5%;   Pred. No. 14;
Matches   26;   Conservative   0;   Mismatches   8;   Indels   0;   Gaps   0;

Db   189989 CAGCATGAGAATACTATTTTCTCACACTAAGGTG 190022

Qy   1 CAGATCTAGATACTAGTTCTACACTCTTACGTG 34
        ||| |||||| ||||||| |||| ||| |||
Db   189989 CAGCATGAGAATACTATTTTCTCACACTAAGGTG 190022

ACCESSION   AL031686
LOCUS   HS981L23/c   94817 bp   DNA   PRI   12-MAY-1999
DEFINITION Human DNA sequence from clone 981L23 on chromosome 20q12.1-13.2.
            Contains a Krueppel type zinc-finger protein pseudogene, a ZNF127
            pseudogene and a KRAB box type zinc-finger protein pseudogene.
            Contains ESTs, an STS ,GSSs and a putative CpG island, complete
            sequence.
ACCESSION   AL031686
VERSION   AL031686.2 GI:4581428
NID   g4581428
KEYWORDS   HTG; CpG island; KRAB box; Krueppel; zinc-finger; ZNF127.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 94817)
  AUTHORS   Babbage,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-MAR-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequests@sanger.ac.uk
COMMENT   On Apr 12, 1999 this sequence version replaced gi:4376020.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 981L23. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            981L23 is from the library RPCI5 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES   Location/Qualifiers
     source   1..94817
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="981L23"
                     /map="q12.1-13.2"
                     /clone_lib="RPCI5"
     repeat_region   83..133
                     /note="L2 repeat: matches 2691..2741 of consensus"
     repeat_region   145..444
                     /note="AluSq repeat: matches 1..300 of consensus"
     misc_feature   complement(537..878)

repeat_region	814..960	/note="match: STS 294362"
repeat_region	961..1258	/note="MIR repeat: matches 51..206 of consensus"
repeat_region	1259..1311	/note="MER33 repeat: matches 1..320 of consensus"
repeat_region	1384..1948	/note="MIR repeat: matches 206..256 of consensus"
misc_feature	1948..3122	/note="L2 repeat: matches 2125..2750 of consensus"
repeat_region	1949..2251	/note="putative CpG island"
repeat_region	2252..2390	/note="AluSx repeat: matches 1..301 of consensus"
repeat_region	2526..2655	/note="MIR repeat: matches 1998..2125 of consensus"
repeat_region	3599..3797	/note="65 copies 2 mer cc 59% conserved"
repeat_region	3811..3899	/note="MIR repeat: matches 16..227 of consensus"
repeat_region	3929..4076	/note="MIR repeat: matches 98..184 of consensus"
repeat_region	4284..4374	/note="MIR repeat: matches 67..212 of consensus"
repeat_region	4988..5172	/note="MER96 repeat: matches 1..93 of consensus"
repeat_region	5260..5557	/note="MIR repeat: matches 68..258 of consensus"
repeat_region	5597..6080	/note="AluSx repeat: matches 2..299 of consensus"
repeat_region	6265..6571	/note="L2 repeat: matches 2219..2728 of consensus"
repeat_region	6637..6799	/note="AluSx repeat: matches 1..299 of consensus"
repeat_region	6894..7117	/note="MIR repeat: matches 85..256 of consensus"
repeat_region	7523..7626	/note="L2 repeat: matches 2068..2412 of consensus"
repeat_region	7663..7964	/note="MIR repeat: matches 20..251 of consensus"
repeat_region	8081..8269	/note="52 copies 2 mer gt 72% conserved"
repeat_region	9420..9771	/note="MLT1A1 repeat: matches 22..365 of consensus"
repeat_region	9748..10112	/note="MIR repeat: matches 21..216 of consensus"
repeat_region	10521..10864	/note="L2 repeat: matches 2383..2710 of consensus"
repeat_region	11136..11313	/note="MIR repeat: matches 118..231 of consensus"
repeat_region	11467..11659	/note="LTR41 repeat: matches 604..766 of consensus"
repeat_region	12018..12052	/note="MIR repeat: matches 1..187 of consensus"
repeat_region	12146..12269	/note="MIR repeat: matches 109..143 of consensus"
repeat_region	12658..12788	/note="MIR repeat: matches 118..231 of consensus"
repeat_region	12827..12855	/note="MIR repeat: matches 107..259 of consensus"
repeat_region	13169..13204	/note="MIR repeat: matches 69..97 of consensus"
repeat_region	13826..14398	/note="18 copies 2 mer tg 97% conserved"
repeat_region	14399..14620	/note="HAL1 repeat: matches 211..756 of consensus"
repeat_region	14621..14925	/note="MER57A repeat: matches 1..211 of consensus"
repeat_region	14926..15117	/note="AluYb8 repeat: matches 1..311 of consensus"
repeat_region	15118..15199	/note="MER57A repeat: matches 211..434 of consensus"
repeat_region	15226..15414	/note="HAL1 repeat: matches 756..840 of consensus"

repeat_region	15226..15414	/note="MER20 repeat: matches 7..213 of consensus"
repeat_region	15771..16335	/note="L1MEc repeat: matches 2372..2612 of consensus"
repeat_region	16355..16478	/note="L1M4 repeat: matches 4940..5051 of consensus"
repeat_region	16386..16481	/note="3 copies 32 mer tt 69% conserved"
repeat_region	16486..16756	/note="AluSg repeat: matches 1..269 of consensus"
repeat_region	16948..17027	/note="40 copies 2 mer ta 65% conserved"
repeat_region	17120..17608	/note="AluSx repeat: matches 1..296 of consensus"
repeat_region	17722..17833	/note="L1M2 repeat: matches 5318..5825 of consensus"
repeat_region	17834..18109	/note="L1M4 repeat: matches 4940..5051 of consensus"
repeat_region	18110..18407	/note="MER33 repeat: matches 1..280 of consensus"
repeat_region	18408..18458	/note="AluJb repeat: matches 1..295 of consensus"
repeat_region	18459..18496	/note="MER33 repeat: matches 280..323 of consensus"
repeat_region	18974..19254	/note="L1M4 repeat: matches 4903..4940 of consensus"
repeat_region	20328..20596	/note="AluSx repeat: matches 1..296 of consensus"
repeat_region	20938..20975	/note="HERVL repeat: matches 5075..5345 of consensus"
repeat_region	20979..21021	/note="19 copies 2 mer gt 95% conserved"
repeat_region	21030..21259	/note="MER5A repeat: matches 26..68 of consensus"
repeat_region	21261..21351	/note="AluSx repeat: matches 1..290 of consensus"
repeat_region	21493..22181	/note="L1MA2 repeat: matches 5102..6308 of consensus"
repeat_region	22404..22828	/note="MER5A repeat: matches 66..179 of consensus"
repeat_region	22833..23063	/note="TIGGER1 repeat: matches 473..2418 of consensus"
repeat_region	23220..23689	/note="L1MC4 repeat: matches 7741..7973 of consensus"
repeat_region	23690..24868	/note="TIGGER1 repeat: matches 1..495 of consensus"
repeat_region	24876..26802	/note="65 copies 2 mer aa 61% conserved"
repeat_region	29065..29165	/note="ML1J repeat: matches 403..516 of consensus"
repeat_region	29181..29282	/note="HERVL repeat: matches 5211..5316 of consensus"
repeat_region	29358..29703	/note="HERV16 repeat: matches 2758..3096 of consensus"
repeat_region	29704..30328	/note="MLT2D repeat: matches 1..553 of consensus"
repeat_region	30329..31860	/note="63 copies 2 mer ac 80% conserved"
repeat_region	32014..32316	/note="HERV16 repeat: matches 1688..2758 of consensus"
repeat_region	32346..32660	/note="AluSx repeat: matches 1..303 of consensus"
repeat_region	33110..33219	/note="MER34 repeat: matches 3..517 of consensus"
repeat_region	33225..33405	/note="L1ME1 repeat: matches 5563..5662 of consensus"

RESULT 8

Qy   4 ATCTAGAATACTATGTTCTACACTCTTACGTGC 35
        ||·||||||| ||||||||||||||||||| |||||
Db 70008 ATTTAGAATAATATGTTCTCCATTTTAAATGC 69977

Query Match   59.4%;   Score 20.8; DB 10;   Length 94817;
Best Local Similarity 78.1%;   Pred. No. 23;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

LOCUS   AF049125   1790 bp   mRNA   ROD   06-MAR-1998
DEFINITION Mus musculus talpoxin-associated calcium binding protein 49 mRNA,
            complete cds.
ACCESSION   AF049125
NID   92935465
VERSION   AF049125.1 GI:2935465
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1790)
  AUTHORS   Dodds,D., Schlimgen,A.K., Lu,S.Y. and Perin,M.S.
  TITLE   Novel reticular calcium binding protein is purified on talpoxin
            columns
  JOURNAL   J. Neurochem. 64 (5), 2339-2344 (1995)
  MEDLINE   95239201
REFERENCE   2 (bases 1 to 1790)
  AUTHORS   Perin,M.S.
  TITLE   Mouse talpoxin-associated calcium binding protein 49 (TCBP49)
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 1790)
  AUTHORS   Perin,M.S.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Neurosciences, Cleveland Clinic Foundation,
            9500 Euclid Avenue, Cleveland, OH 44195, USA
FEATURES   Location/Qualifiers
     source   1..1790
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS   194..1159
                     /codon_start=1
                     /note="TCBP49; contains 6 EF hand domains"
                     /product="talpoxin-associated calcium binding protein 49"
                     /protein_id="AAC05132.1"
                     /db_xref="PID:g2935466"
                     /db_xref="GI:2935466"
                     /translation="MRLGRPPAALGLLPLLLYAAVAGASKAEELHYPQGEHRADYDR
                     EALLGVOEDVDEYVKLGHEEQORRLQSIIKKIDSDSDGFIENELSQWIOMSFKHYAM
                     QEAKQOFVEIDKNSDGDPVTWDEYNIQMYDRVIDFDENTALDDTEGGSFRQLHLKDKK
                     RFEKANODSGPGLSLEEFIAFEHPEEVDYMTEFVIQEALEEHDKNGDGFVSLEEFLGD
                     YRRDPTANEDDEWILVEKDRRVNDYDKDNGRLDPQELLSWVVPNNOGIAQEEALHLI
                     DEMDLNSDKKLSEEELLENQOLFLTSDQTDYGRQLHDDYFHDEL"
BASE COUNT   490 a   384 c   449 g   467 t
ORIGIN

RESULT 9

Qy   2 AGAGTCTAGATACTATGTTCTACACTCTTACGT 33
        |||·||||| |||||||||||||||||||||| ||
Db 1226 AGATATAGAGATGTGTCTTTACACTCTTAAGT 1257

Query Match   59.4%;   Score 20.8; DB 12;   Length 1790;
Best Local Similarity 78.1%;   Pred. No. 43;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

LOCUS   G04095   235 bp   DNA   STS   19-OCT-1995
DEFINITION human STS WI-2964.
ACCESSION   G04095
NID   g721053
VERSION   G04095.1 GI:721053
KEYWORDS   STS sequence; primer; sequence tagged site.
SOURCE   human Random genome wide STSs created from sheared whole human DNA.
  ORGANISM Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
            Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 235)
  AUTHORS   Hudson,T.
  TITLE   Whitehead Institute/MIT Center for Genome Research; Random Genome
            Wide STSs
  JOURNAL   Unpublished (1995)
REFERENCE   2 (bases 1 to 235)
  AUTHORS   Hudson,T.
  TITLE   Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT   Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu
            Primer A: GAGGATGGCACCCCTCTC
            Primer B: AATGCTCCCTGTGTGGAATG
            STS size: 181
            PCR Profile:
              Presoak:
              Denaturation:
              Annealing: 56 degrees C
              Polymerization:
              PCR Cycles: 35
              Thermal Cycler:
            Protocol:
              Template: 10 ng
              Primer: each 5 pM
              dNTPs: each 4 mM
              Taq Polymerase: 0.025 units/ul
              Total Vol: 20 ul
            Buffer:
              MgCl2: 1.5 mM
              KCl: 50 mM
              Tris-HCL: 10 mM
              pH: 9.3.
FEATURES   Location/Qualifiers
     source   1..235
                     /organism="Homo sapiens"
     STS   1..235
     primer_bind   55..72
     primer_bind   complement(217..235)
BASE COUNT   65 a   59 c   61 g   47 t   3 others
ORIGIN

RESULT 10

Qy   1 CAGATCTAGATACTATGTTCTACACTCTTACGTGC 35
        || |||| |||·|| |||||| | |||| |||| |||
Db   55 CAGATCAAGATGGTCTTTTCTAACCTCTTCAGTGC 21

Query Match   58.9%;   Score 20.6; DB 13;   Length 235;
Best Local Similarity 74.3%;   Pred. No. 72;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

LOCUS   SPCC1902   5701 bp   DNA   PLN   25-MAR-1999
DEFINITION S.pombe chromosome III cosmid c1902.

```
ACCESSION   AL049521
NID         g4539590
VERSION     AL049521.1  GI:4539590
KEYWORDS    gaf1; GATA zinc finger; RNA-binding region; RNP-1 signature.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Archiascomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 5701)
  AUTHORS   Seeger,K., Harris,D., Wood,V., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1999) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            During 1995 to 1996 about 66% of S. pombe chromosome 1 was
            sequenced by the Sanger Centre. The sequencing of the S. pombe
            genome is now being continued with funding from The European
            Commission. Fourteen European sequencing laboratories, including
            the Sanger Centre, are participating in the project.
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program SPsplice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites. CDS are numbered using the following
            system eg SPBC25H2.O1c. SP (S. pombe), B (chromosome 2), c25H2
            (cosmid name), O1 (first CDS), c (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid c1902 is
            overlapped at the 5' end by cosmid c663, EMBL entry SPCC663,
            accession number AL031307 and EMBL entry SPCC417, accession number
            AL035076.
FEATURES             Location/Qualifiers
     source          1..5701
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="III"
                     /clone="cosmid c1902"
                     /map="IIIR"
     gene            1..2184
                     /gene="gaf1"
     CDS             1..2184
                     /gene="gaf1"
                     /partial
                     /codon_start=1
                     /note="SPCC1902.O1, len:727"
                     /gene="gaf1"
                     /label=gaf1
                     /product="gaf1 protein"
                     /protein_id="CAB40003.1"
                     /db_xref="PID:e1421489"
                     /db_xref="GI:4539591"
                     /db_xref="PID:g4539591"
                     /translation="ADTTGSSLMEFNYIQRVRKTSFDESTAKSKKRSIADSHFPDEN
                     AMRPHDLESQOFSYRKIHASNSFNEYVRDIDSSNFSNLDASALPISPPSDSFFSVHSH
                     NLPNAPSIPANSNNSASPNNRIKASPKHADYDVLGLFDFMYPSEPSSFFPENGGFPSF
                     VDANTHEQTLPFSSATNSFSFEHGSGAGFPIPGSSVPSTSYHANTASEDGFSSSYNSOGL
                     FGISSPLSSGVTPNQSSFFPDVSGGNNIFDVSRNNHEVSSPLIQSPGSYVSMPSINMVSS
                     LPISAFVPNSNSQOFPRPRNTRTNSKSKSVGGOSSGSSVGDSNQENAESSNFSSSHNSAEW
                     ASGETTGHSSSNSPLPGSDMFSQOFMRVGTAMGVAPVPVRSNSSNNFGQNFFHQTSPQFSA
                     VPHRKVSAQODTNLMGSSPGMYNHMPILNRATSANSISPGVLPEGMAASLKRTTNTA
                     ATPQAALPTTLDTKKDRSVSFNINKNAEKPTVSNAAEDKKGDANTRRANAINPTPICT
```

```
     misc_feature    complement(1..127)
                     /gene="gaf1"
                     /note="nominal overlap with SPCC417 S. pombe chromosome 3"
     misc_feature    1507..1626
                     /gene="gaf1"
                     /note="Match to PF00320 GATA, GATA zinc finger"
     gene            3696..5599
                     /gene="SPCC1902.02"
     CDS             join(3696..3735,3850..4011,4077..5599)
                     /gene="SPCC1902.02"
                     /note="SPCC1902.02, len:574,
                     SIMILARITY:Schizosaccharomyces pombe, YAI7_SCHPO,
                     hypothetical 62.5 kd protein c24bl1.07c in chromosome 1.,
                     (561 aa), fasta scores: opt: 368, E():9.3e-15, (25.9%
                     identity in 517 aa)"
                     /codon_start=1
                     /label=SPCC1902.02
                     /product="hypothetical protein"
                     /protein_id="CAB40004.1"
                     /db_xref="PID:e1421490"
                     /db_xref="GI:4539592"
                     /db_xref="PID:g4539592"
                     /translation="MSEEDTSKLRTLSVGSNAIAFISWRLSESKACHTILIWNRCE
                     SVLSEGIRRSSVFGSSTKWKPDVVAPTVEOLAMNSEPFDIFVCLKILPSVINLDTAI
                     KEVVTPGHTCIVLNTTGIVGAEKELQHAFPNNPVLSFVLPDFAQRGPLQFEHTTFFA
                     DSAKSVIYGLTEEDDVPDSVQDAMIEDTLTLTLEAGGVSCDFLSKIQKKQWETGVGH
                     MCFYPLSIINDEPNLALMIRLKSFAKVIDGLMDEARFSIAQOAQGCEFEPEKLDVLKRHI
                     VNRMLATRPRSYIPYQDYIAHRPLEVAVLLGIYPVIIAKELGVSVRMETLLALFDAKNK
                     RNLTVRAGTPQOSSNFNPAMRRSPVGAAASRSPSRSTIGISNRIGSVDLLNTRQFTSS
                     PIGGSMPKGPNSIYIKIPSASMVNLSSPLVTSPSGLNPTGRPSREFGGRVRGNPLTMNKA
                     GSVSDLLSTSTNMASSDALETASMIGVPSAMPPNSFDMLTLTQRRNRNRNNQSSSPPAP
                     SDRRYTMGRARRPVQTRGMIDSVIFLEDDMSTLYDTSRYPTRNSKPATPKASRPPSIA
                     STVHRRMD"
     misc_feature    3736..3741
                     /gene="SPCC1902.02"
                     /note="gtatgt, splice donor sequence"
     misc_feature    4012..4017
                     /gene="SPCC1902.02"
                     /note="gtatgg, splice donor sequence"
     misc_feature    3836..3849
                     /gene="SPCC1902.02"
                     /note="ctaatcgagtcatttag, splice branch and acceptor"
     misc_feature    4059..4076
                     /gene="SPCC1902.02"
                     /note="ctaatcgagtcatttag, splice branch and acceptor"
     misc_feature    4516..4539
                     /gene="SPCC1902.02"
                     /note="RNP-1 signature"
     misc_feature    complement(5599..5701)
                     /gene="SPCC1902.02"
                     /note="nominal overlap with SPCC663 S. pombe chromosome 3"
BASE COUNT      1525 a   1180 c   1074 g   1922 t
ORIGIN
```

```
RESULT 11
LOCUS       HS1145L23
DEFINITION  HS1145L23  203006 bp  DNA                     PRI     08-OCT-1999
            Human DNA sequence from clone 1145L23 on chromosome Xp21.1-21.3,
            complete sequence.

Query Match             57.7%;  Score 20.2;  DB 7;  Length 5701;
Best Local Similarity   75.8%;  Pred. No. 65;
Matches   25;  Conservative    0;  Mismatches    8;  Indels   0;  Gaps    0;

Qy      1 CAGATCTAGATACTANTGTTCTACACTCTTACGT 33
          |||| ||| || || |||| | ||| | |||||||| |
Db   4416 CAGTTCAAGATGCAATGATCGAAACTCTTACAT 4448
```

```
ACCESSION   AL096699
NID         g6015529
VERSION     AL096699.11  GI:6015529
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 203006)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 7, 1999 this sequence version replaced gi:5931812.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is
            the entire insert of clone 1145L23. This sequence has been finished
            to resolve all sequencing problems, such as compressions and
            repeats, but not necessarily within known annotated human repeat
            sequence elements (e.g. Alu). Where the sequence is ambiguous,
            there is an annotation using the 'unsure' feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the Sanger Centre Chromosome X
            Mapping Group. Further Information can be found at
            http://www.sanger.ac.uk/HGP/Chrx
            1145L23 is from the library RPCI-5 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..203006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="1145L23"
                     /map="p21.1-21.3"
                     /clone_lib="RPCI-5"
BASE COUNT    67748 a  37548 c  35939 g  61771 t
ORIGIN

Qy      1 CAGAGTCTAGAGTAGTTCTACACTCTTACGT 33
          ||| ||| |||||||| ||| |||||| ||||||
Db 106668 CAGCTCAAGAGCCTAGTCCTCTTTACTCTTACAT 106700

Query Match     57.7%;  Score 20.2;  DB 10;  Length 203006;
Best Local similarity   75.8%;  Pred. No. 37;
Matches   25;  Conservative   8;  Mismatches   8;  Indels   0;  Gaps   0;


RESULT 12
AC002476/c
LOCUS       AC002476   147102 bp   DNA        PRI   21-AUG-1997
DEFINITION  Human PAC clone DJ318C15 from Xq23, complete sequence.
ACCESSION   AC002476
NID         g2340101
VERSION     AC002476.1  GI:2340101
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 147102)
  AUTHORS   Graves,T and Ozersky,P.
  TITLE     The sequence of H. sapiens PAC clone DJ318C15
  JOURNAL   Unpublished (1997)
REFERENCE   2 (bases 1 to 147102)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:saplens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by David Bentley's chromosome X
            mapping group at the Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, UK.  Further Information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            VECTOR: pCYPAC2

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1 prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/  The clone is available
            from Genome Systems, Inc. (http://www.genomesystems.com).
            VECTOR: pCYPAC2

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of H_DJ318C15;
            actual end is at 147102 of H_DJ318C15.  The orientation of this
            clone is unknown.
FEATURES             Location/Qualifiers
     source          1..147102
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="DJ318C15"
                     /clone_lib="RPCI-1"
                     /map="Xq23"
     misc_feature    1..332
                     This clone contains STS's WI-13810 (NID:g1343514) and WI-9475
                     (NID:g860548).
     misc_feature    1..332
                     /note="match to EST N24741 (NID:g1388891) yx56c02.s1"
     misc_feature    1..440
                     /note="match to EST N31640 (NID:g1152039) yw84903.s1"
     misc_feature    1..569
                     /note="match to EST AA207254 (NID:g1802747) zq82e06.s1"
     misc_feature    complement(14..401)
                     /note="match to EST H23858 (NID:g892553) yn71f07.r1"
     misc_feature    complement(295..663)
                     /note="match to EST H33858 (NID:g9486641)"
     misc_feature    complement(392..499)
                     /note="match to EST R23654 (NID:g778542) yh35g02.r1"
     misc_feature    395..811
                     /note="match to EST N80157 (NID:g1242858) za65e08.s1"
```

```
misc_feature    complement(432..753)
                /note="match to EST R23654 (NID:g778542) yh35g02.r1"
misc_feature    complement(481..958)
                /note="match to EST N42209 (NID:g1166240) yw71e02.r1"
misc_feature    complement(692..1162)
                /note="match to EST AA209339 (NID:g1807353) zq82e06.r1"
misc_feature    complement(1269..1393)
                /note="match to EST H09881 (NID:g874703) ym05b04.r1"
misc_feature    complement(1336..1393)
                /note="match to EST H09881 (NID:g874703) ym05b04.r1"
misc_feature    complement(1418..1676)
                /note="match to EST R25050 (NID:g779938) yg42a06.r1"
misc_feature    complement(1418..1676)
                /note="match to EST R25050 (NID:g779938) yg42a06.r1"
misc_feature    complement(1638..1676)
                /note="match to EST H09881 (NID:g874703) ym05b04.r1"
repeat_region   1677..1716
                /rpt_family="L1"
misc_feature    complement(1717..1972)
                /note="match to EST AA148755 (NID:g1718917) zo09c05.r1"
misc_feature    complement(1717..1972)
                /note="match to EST AA148755 (NID:g1718917) zo09c05.r1"
repeat_region   complement(2097..2120)
                /rpt_family="L1"
repeat_region   2478..2513
                /rpt_family="L1"
repeat_region   2894..3036
                /rpt_family="L1"
repeat_region   complement(2897..3017)
                /rpt_family="L1"
misc_feature    4778..5126
                /note="match to EST R31677 (NID:g787520) yh63c10.s1"
misc_feature    4779..5302
                /note="match to EST AA085744 (NID:g1629216) zk71b09.s1"
misc_feature    4779..4982
                /note="match to EST F01897 (NID:g645454) ..."
misc_feature    complement(4779..5124)
                /note="match to EST R65700 (NID:g838338) yi16f12.r1"
misc_feature    complement(4779..4878)
                /note="match to EST C21402 (NID:g1625212)"
misc_feature    complement(4790..4985)
                /note="match to EST D58313 (NID:g964935)"
misc_feature    complement(5072..5302)
                /note="match to EST H48596 (NID:g988436) yq80a02.r1"
misc_feature    5100..5194
                /note="match to EST R31677 (NID:g787520) yh63c10.s1"
misc_feature    5101..5443
                /note="match to EST R97806 (NID:g983466) yq60d03.s1"
misc_feature    complement(5132..5302)
                /note="match to EST AA125797 (NID:g1687768) z122d02.r1"
gene            complement(5161..43007)
                /gene="LAMP2"
CDS             complement(join(5161..5300,15568..15732,16437..16500,
                20143..20225,21679..21863,22808..22966,29195,29408,
                30489..30607,42944..43007))
                /gene="LAMP2"
                /codon_start=1
                /product="lysosome-associated membrane protein-2 (LAMP2)"
                /protein_id="AAB67314.1"
                /db_xref="PID:g2340103"
                /db_xref="GI:2340103"
                /translation="MVCFRLFPVPGSGLVLVCLVLGAVRSYALELNLTDSENATCLYA
                KWQMNFTVRYETTNKTIYKTVTISDHGTVTYNGSSICGDDQNGPKIAVQFGPGFSWIANF
                TKAASTYSIDSVSFSYNTGDNTTPDAEDKGILTVDELLAIRIPLNDLFRCNSLSTLE
                KNDVVQHYWDVLVQAFVQNGTVSTNEFLCDKDKTSTVAPTIHTTVPSPTTTPTPKEKP
                EAGTYSVNNGNDTCLLATMGLQLNITQDKVASVININPNTHSTGSCRSHTALLRINS
                STIKLDFVFAVKNENRFYLKEVNISMYLVNGSVFSIANNLSYWDAPLGSSYMCNKE
                QTVSVSGAFQINTFDFAVQPQPFNTGKYSTAQDCSADDNFLVPIAVGAALAGVLILV
                LLAYFIGLKHHHAGYEQF"
misc_feature    complement(5538..6010)
                /note="match to EST R97855 (NID:g983515) yq60d03.r1"
```

```
misc_feature    complement(5818..6167)
                /note="match to EST H22728 (NID:g891423) ym49e06.r1"
                /gene="LAMP2"
repeat_region   6649..6782
                /rpt_family="ALU"
                /gene="LAMP2"
misc_feature    complement(7360..7421)
                /note="match to EST N63463 (NID:g1211292) yy61d02.s1"
                /gene="LAMP2"
repeat_region   8685..8981
                /note="match to EST R97855 (NID:g983515) yq60d03.r1"
                /rpt_family="ALU"
                /gene="LAMP2"
repeat_region   complement(10334..10498)
                /rpt_family="ALU"
                /gene="LAMP2"
misc_feature    10335..10766
                /note="match to EST D29162 (NID:g599095)"
                /gene="LAMP2"
misc_feature    10338..10452
                /note="match to EST N63482 (NID:g1211311) yy61g05.s1"
                /gene="LAMP2"
misc_feature    complement(10486..10881)
                /note="match to EST AA052944 (NID:g1544144) z170c10.s1"
                /gene="LAMP2"
misc_feature    complement(10509..11005)
                /note="match to EST N77754 (NID:g1240455) yz83b04.r1"
                /gene="LAMP2"
misc_feature    complement(10660..10793)
                /note="match to EST D62941 (NID:g966715)"
                /gene="LAMP2"
misc_feature    complement(10684..11028)
                /note="match to EST R78159 (NID:g853269) yi81b06.r1"
                /gene="LAMP2"
misc_feature    complement(10733..11121)
                /note="match to EST R78159 (NID:g853269) yi81b06.r1"
                /gene="LAMP2"
misc_feature    complement(10900..11121)
                /note="match to EST N94781 (NID:g1267060) yy61g05.r1"
                /gene="LAMP2"
misc_feature    complement(11152..11249)
                /note="match to EST AA043891 (NID:g1521749) zk57b02.r1"
                /gene="LAMP2"
misc_feature    complement(11275..11341)
                /note="match to EST AA043891 (NID:g1521749) zk57b02.r1"
                /gene="LAMP2"
```

```
RESULT  13
HSU60822/c
LOCUS       HSU60822      112359 bp    DNA             PRI       23-JUL-1997
DEFINITION  Human dystrophin (DMD) gene, exons 7, 8 and 9, and partial cds.
ACCESSION   U60822 L08092 U06836
NID         g2275172
VERSION     U60822.1  GI:2275172
KEYWORDS    110kb intron; repetitive sequences; THE-1 sequences; translocation
            breakpoints.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 31798 to 63822)
  AUTHORS   McNaughton,C.C., Broom,J.E., Hill,D.F., Jones,W.A., Marshall,C.J.,
            Renwick,N.M., Stockwell,P.A. and Petersen,G.B.
  TITLE     A cluster of transposon-like repetitive sequences in intron 7 of
            the human dystrophin gene
  JOURNAL   J. Mol. Biol. 232 (1), 314-321 (1993)
  MEDLINE   93323127

Qy       2 AGATCTAGATACTATGTTCTACACACTCTTACGTG 34
           ||||||||| |||||||| |||||||| ||||||||
Db   54603 AGCTCTAGAAAAATACTTCTTCACTCTACCCTG 54571

Query Match             57.7%;  Score 20.2;  DB 11;  Length 147102;
Best Local Similarity   75.8%;
Matches    25;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps   0;
```

```
REFERENCE   2  (bases 92090 to 94459)
  AUTHORS   McNaughton,J.C., Marshall,C.J., Broom,J.E., Hughes,G., Jones,W.A.,
            Stockwell,P.A. and Petersen,G.B.
  TITLE     Phylogenetic relationships among transposon-like elements in human
            and primate DNA
  JOURNAL   J. Mol. Evol. 40 (2), 127-135 (1995)
  MEDLINE   95214090
REFERENCE   3  (bases 1 to 112359)
  AUTHORS   McNaughton,J.C., Hughes,G., Jones,W.A., Stockwell,P.A., Klamut,H.J.
            and Petersen,G.B.
  TITLE     The evolution of an intron: analysis of a long, deletion-prone
            intron in the human dystrophin gene
  JOURNAL   Genomics 40 (2), 294-304 (1997)
  MEDLINE   97237049
REFERENCE   4  (bases 1 to 112359)
  AUTHORS   McNaughton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1996) Department of Biochemistry, University of
            Otago, Box 56, Dunedin, New Zealand
REFERENCE   5  (bases 1 to 112359)
  AUTHORS   McNaughton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1997) Department of Biochemistry, University of
            Otago, Box 56, Dunedin, New Zealand
  REMARK    Sequence update by submitter
COMMENT     On Jul 23, 1997 this sequence version replaced gi:1845599.
FEATURES             Location/Qualifiers
     source          1..112359
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp21.2"
                     /clone="cosmids XJC5, Tcos5, P31"
     repeat_region   5..223
                     /gene="DMD"
                     /rpt_family="L1"
     Intron          <37..1024
                     /gene="DMD"
                     /number=6
     gene            complement(442..518)
                     /gene="DMD"
                     /rpt_family="L1MA9"
     repeat_region   complement(523..832)
                     /gene="DMD"
                     /rpt_family="Alu-Jb"
     exon            1025..1143
                     /gene="DMD"
                     /number=7
     CDS             join(<1025..1143,110718..110899,112015..>112143)
                     /gene="DMD"
                     /codon_start=2
                     /product="dystrophin"
                     /protein_id="AAC51631.1"
                     /db_xref="PID:g1845600"
                     /db_xref="GI:1845600"
                     /translation="PDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDVDTT
                     YPDKKSILMYITSLFGVLPQQVSIEAIQEVEMLRPPKVTKEEHFQLHHQMHYSQOIT
                     VSLAQGYERTSSPKPRFKSYRAYTQAAYVTISDPTRSPFPSQ"
     Intron          1144..110717
                     /gene="DMD"
                     /number=7
     repeat_region   complement(2901..2941)
                     /gene="DMD"
                     /rpt_family="L1MA2"
     repeat_region   complement(3288..6448)
                     /gene="DMD"
                     /rpt_family="L1"
     repeat_region   6449..6544
                     /gene="DMD"
                     /rpt_family="L1"
     repeat_region   6545..10074
                     /gene="DMD"
                     /rpt_family="L1"

     misc_feature    10324
                     /gene="DMD"
                     /note="deleted in sequence with EMBL accession number
                     221689 in association with t(X;1)"
     misc_feature    10606..10648
                     /gene="DMD"
                     /note="t(X;1) breakpoint noted in sequence with EMBL
                     accession number 221689"
     repeat_region   complement(13351..13495)
                     /gene="DMD"
                     /rpt_family="L1MA5"
     repeat_region   complement(13886..16015)
                     /gene="DMD"
                     /rpt_family="Alu-J (FRAM)"
     repeat_region   complement(16016..16327)
                     /gene="DMD"
                     /rpt_family="L1MA5"
     repeat_region   complement(16328..19040)
                     /gene="DMD"
                     /rpt_family="Alu-Ya"
     repeat_region   19041..19302
                     /gene="DMD"
                     /rpt_family="L1"
     repeat_region   complement(22862..23161)
                     /gene="DMD"
                     /rpt_family="MER44"
     satellite       23349..23422
                     /gene="DMD"
                     /rpt_family="MER44"
     satellite       24514..24547
                     /gene="DMD"
     repeat_region   complement(26830..27016)
                     /gene="DMD"
                     /rpt_family="L1PA15"
     satellite       27230..27689
                     /gene="DMD"
                     /rpt_family="L1PA15"
     repeat_region   complement(27017..27143)
                     /gene="DMD"
                     /rpt_family="MER2"
     repeat_region   complement(27690..27945)
                     /gene="DMD"
                     /rpt_family="L1MA2"
     repeat_region   complement(29491..29791)
                     /gene="DMD"
                     /rpt_family="Alu-Ya"
     repeat_region   30071..30119
                     /gene="DMD"
                     /rpt_family="THE-1"
     repeat_region   complement(30756..30916)
                     /gene="DMD"
                     /rpt_family="L1"
     repeat_region   30942..31113
                     /gene="DMD"
                     /rpt_family="L1"
     repeat_region   31689..32181
                     /gene="DMD"
                     /rpt_family="MLT1D"
     repeat_region   33905..36231
                     /gene="DMD"
                     /rpt_family="THE-1"
     repeat_region   36934..37375
                     /gene="DMD"
                     /rpt_family="L1MA5"
     repeat_region   38036..38212
                     /gene="DMD"
                     /rpt_family="MIR"
     satellite       39126..39147
                     /gene="DMD"
     repeat_region   41174..41514
                     /gene="DMD"
                     /rpt_family="THE-1 (LTR)"
     repeat_region   complement(43736..43934)
                     /gene="DMD"
                     /rpt_family="L1MA5"
     repeat_region   44014..45070
                     /gene="DMD"
                     /rpt_family="L1"
```

Left column (RESULT 14)

```
repeat_region    /rpt_family="L1"
                 complement(45071..45483)
                 /gene="DMD"
repeat_region    /rpt_family="THE-1 (LTR)"
                 45484..46342
                 /gene="DMD"
repeat_region    /rpt_family="L1"
                 complement(48450..48777)
                 /gene="DMD"
repeat_region    /rpt_family="Alu-Sc"
                 complement(50028..50145)
                 /gene="DMD"
repeat_region    /rpt_family="MIR2"
                 complement(50145..
                 /gene="DMD"
repeat_region    /rpt_family="MIR"
                 complement(51861..51959)
                 /gene="DMD"
satellite        5678..52692
                 /gene="DMD"
                 /note="(TCC)15"
repeat_region    complement(52837..53199)
                 /gene="DMD"
                 /rpt_family="THE-1 (LTR)"
misc_feature     53051..54002
                 /gene="DMD"
                 /note="inverted in sequence with GenBank accession number
                 M18024 in association with t(X;21)(p21;p12)"
repeat_region    53342..53583
                 /gene="DMD"
                 /rpt_family="L1MA9"
misc_feature     55930
                 /gene="DMD"
                 /note="t(x;21)(p21;p12) breakpoint noted in sequence with
                 GenBank accession number M18024"
repeat_region    56250..56698
                 /gene="DMD"
                 /rpt_family="MIR"
```

```
Query Match
Best Local Similarity   57.7%;   Score 20.2;   DB 11;   Length 112359;
Matches   25;   Conservative   0;   Mismatches   8;   Indels   0;   Gaps   0;

Qy     1 CAGATCTRAGAACTAAGTTCTACACTCTTACGT 33
         ||| || |||| |||||||||||||| |||||||| |
Db 38060 CAGCTCAAGAGCCTATCGTCTTTACTCTTACAT 38028
```

```
RESULT 14
LOCUS       G21425/c
G21425/c
LOCUS       G21425          439 bp    DNA        STS    31-MAY-1996
DEFINITION  human STS WI-13400, sequence tagged site.
ACCESSION   G21425
NID         g1341751
VERSION     G21425.1  GI:1341751
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human STS derived from sequences in dbEST and the Unigene
            collection.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 439)
AUTHORS     Hudson,T.
TITLE       Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
JOURNAL     Unpublished (1995)
COMMENT     Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu
```

Right column (RESULT 15)

```
Primer A: TGTTTGTATTCACTTCAJGTAGCA
Primer B: AAAAGGTTCAAGGAATTTATCTCA
STS size: 129
PCR Profile:
  Presoak:
  Denaturation:
  Annealing: 56 degrees C
  Polymerization:
  PCR Cycles: 35
Thermal Cycler:
Protocol:
  Template: 10 ng
  Primer: each 5 pM
  dNTPs: each 4 nM
  Taq Polymerase: 0.025 units/ul
  Total Vol: 20 ul
Buffer:
  MgCl2: 1.5 mM
  KCl: 50 mM
  Tris-HCL: 10 mM
  pH: 9.3
```

```
Derived from dbEST (genbank accession R49688).

FEATURES          Location/Qualifiers
source            1..439
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /map="403.0 cR from top of Chr12 linkage group"
STS               44..172
primer_bind       44..68
primer_bind       complement(148..172)
BASE COUNT    135 a    67 c    82 g    149 t    6 others
ORIGIN
```

```
Query Match
Best Local Similarity   57.7%;   Score 20.2;   DB 13;   Length 439;
Matches   25;   Conservative   0;   Mismatches   9;   Indels   0;   Gaps   0;

Qy   1 CAGATCTAGATAACTATATGTTCTACACTCTTACGTG 34
       || ||| ||| |||| ||||||||| ||||||  || ||
Db 269 CATATGTATATAGTATGTTGTTACACTGNCACATG 236
```

```
RESULT 15
LOCUS       AC006962/c
AC006962/c
LOCUS       AC006962      163384 bp    DNA        HTG    05-MAR-1999
DEFINITION  Homo sapiens clone DA0052J04, WORKING DRAFT SEQUENCE, 3 unordered
            pieces.
ACCESSION   AC006962
NID         g4337300
VERSION     AC006962.1  GI:4337300
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 163384)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 163384)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (05-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
 *  This record will be updated with the finished sequence
 *  as soon as it is available and the accession number will
 *  be preserved.
 *
 *      1      12837:  contig of 12837 bp in length
 *  12838      12855:  gap of unknown length
 *  12856      75375:  contig of 62520 bp in length
 *  75376      75393:  gap of unknown length
 *  75394     163384:  contig of 87991 bp in length.
FEATURES          Location/Qualifiers
     source       1..163384
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="DA0052J04"
BASE COUNT    50381 a   33116 c   32647 g   47183 t    57 others
ORIGIN
```

```
Query Match      57.7%;    Score 20.2;  DB 34;   Length 163384;
Best Local Similarity   75.8%;    Pred. No. 38;
Matches   25;  Conservative   0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 AGATCTAGATACTATGTTCTACACTCTTACGTG 34
           ||·||||||·|·|||·|||||·||||||·||·||
Db  131684 AGCTCTAGAAAATATCTTCTTCACTCTACCCTG 131652
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 1998 Compugen Ltd.
GenCore version 4.5

Run on: December 29, 1999, 17:02:53 ; Search time 12.66 Seconds
                 (without alignments)
                 1359.937 Million cell updates/sec

Title: US-09-077-173A-2
Perfect score: 1946
Sequence: 1 MASTESSLLRSLGLSPGPGS..........CRWAATPQDSSCSTPRADRF 365

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database :   PIR_62:*

Word size :   0

Number of hits that pass the threshold :   142080

                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Length	DB	ID	Description
1	1940	99.7	365	2	S68679	G protein-coupled
2	962.5	49.5	373	2	JC2134	angiotensin II rec
3	934	48.0	375	2	JN0763	somatostatin recep
4	641.5	33.0	373	2	JC4737	P-2U nucleotide re
5	630	32.4	372	2	S54253	P2Y1 receptor G-pr
6	628	32.3	362	2	S33733	purinergic recepto
7	621	31.9	373	2	JC4162	G protein-coupled
8	588	30.2	328	2	S55450	P2Y receptor - bov
9	588	30.2	326	2	G02514	a novel G protein-
10	586	30.1	328	2	JC4800	P2 purinoceptor -
11	501.5	25.8	308	2	I50241	P2y6 receptor - hu
12	479.5	24.6	344	2	T09508	G protein-coupled
13	450.5	23.2	370	2	JC5549	intron 17 purinerg
14	423	21.7	399	2	I48705	heptahelical P2Y5-
15	411.5	21.1	420	2	I11667	proteinase activat
16	403	20.7	397	2	S66518	thrombin receptor
17	403	20.7	397	2	S64709	proteinase-activat
18	403	20.7	369	2	G02131	proteinase-activat
19	400	20.6	372	2	I38532	delta opioid recep
20	397.5	20.4	425	2	A37912	thrombin receptor
21	388.5	20.0	427	2	S17148	alpha-thrombin rec
22	388	19.9	432	2	A43448	thrombin receptor
23	385	19.8	372	2	S34592	delta opioid recep
24	384	19.7	361	2	I38657	delta opiate recep
25	383	19.7	361	2	B45680	G protein-coupled
26	382.5	19.7	362	2	JN0694	angiotensin II rec
27	381.5	19.6	398	2	I56504	angiotensin II rec
28	378.5	19.5	400	2	I56553	opiate receptor
29	378.5	19.5	392	2	S65693	opioid receptor mu
30	377	19.4	372	2	B48227	delta opioid recep
31	376.5	19.4	398	2	I56517	mu-opioid recep
32	371.5	19.3	342	2	A40191	platelet-activatin
33	371.5	19.1	398	2	A57510	mu opioid receptor
34	370.5	19.1	359	2	JC5277	G protein-coupled
35	369.5	19.0	359	2	JH0621	angiotensin II rec
36	365	18.8	363	2	I57955	somatostatin recep
37	364.5	18.7	359	2	JC2134	angiotensin II rec
38	364	18.7	364	2	JN0763	somatostatin recep
39	363	18.6	359	2	A42656	angiotensin II rec
40	362	18.6	354	2	S68207	G protein-coupled
41	361	18.5	354	2	I53033	G protein-coupled
42	360.5	18.5	359	2	S44425	angiotensin II rec
43	360	18.5	359	2	JC1104	angiotensin II rec
44	359	18.4	362	2	A57641	G protein-coupled
45	358.5	18.4	359	2	S15403	angiotensin II rec

ALIGNMENTS

RESULT   1
S68679
G protein-coupled receptor - human
C;Species: Homo sapiens (man)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 24-Sep-1999
C;Accession: S68679
R;Stam, N.J.; Klomp, J.; van de Heuvel, M.; Olijve, W.
FEBS Lett. 384, 260-264, 1996
A;Title: Molecular cloning and characterization of a novel orphan receptor (P(2P)) ex
A;Reference number: S68679; MUID:9619780l
A;Accession: S68679
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-365 <STA>
A;Cross-references: EMBL:X96597; NID:g1296631; PIDN:CAA65415.1; PID:e229605; PID:g129
A;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor

Query Match   99.7%; Score 1940; DB 2; Length 365;
Best Local Similarity   100.0%; Pred. No. 3.8e-164;
Matches 364; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 MASTESSLLRSLGLSPGPGSSSEVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLF 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MASTESSLLRSLGLSPGPGSSSEVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLF 60

Qy   61 IFRLRPWDATATYMFHLALSDTLYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 IFRLRPWDATATYMFHLALSDTLYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLY 120

Qy  121 CSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVWLVVAGCLVPNLFVTTSNKG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVWLVVAGCLVPNLFVTTSNKG 180

Qy  181 TTVLCHDTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGCSAQSSSR 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTVLCHDTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGCSAQSSSR 240

Qy  241 LRSLRTIAVVLTVFAVCFVPFHITRTIYYLARLLEADCRVLNIVNVVKVTRPLASANSC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LRSLRTIAVVLTVFAVCFVPFHITRTIYYLARLLEADCRVLNIVNVVKVTRPLASANSC 300

Qy  301 LDPVLYLLLTGDKYRRQLRQLCGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCSTP 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 LDPVLYLLLTGDKYRRQLRQLCGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCSTP 360

Qy  361 RADR 364
        ||||
Db  361 RADR 364
```

RESULT   2
A47556
ATP receptor P2u - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
C;Accession: A47556

R;Lustig, K.D.; Shiau, A.K.; Brake, A.J.; Julius, D.
Proc. Natl. Acad. Sci. U.S.A. 90, 5113-5117, 1993
A;Title: Expression cloning of an ATP receptor from mouse neuroblastoma cells.
A;Reference number: A47556; MUID:93281707
A;Accession: A47556
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-373 <LUS>
A;Superfamily: ATP receptor P2u
C;Cross-references: GB:L14751; NID:g309457; PIDN:AAA39871.1; PID:g309458
C;Keywords: transmembrane protein

Query Match   49.5%; Score 962.5; DB 2; Length 373;
Best Local Similarity 53.7%; Pred. No. 9.7e-78;
Matches 188; Conservative 46; Mismatches 97; Indels 19; Gaps 3;

```
Qy   22 EVELDCWFDEDKFPILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDKATATYMFHLALSD  81
         |: |||||||||||||| |||||||:||||| ||||| ||||:|||:||||
Db   20 ELGYKCRFNEDKFXVLLPVSYGVVCVLGICLNAVGLYIFLCRLKTWNASTTTMFHLAVSD  79

Qy   82 TLYVLSLPTLIIYYAAHHNWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPL 141
         :||||| ||||||||||| |:||| |||| ||||||||||:|||||:|| ||||| ||
Db   80 SLYAASLPLLVIYYARGDHWPFSTVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPL 139

Qy  142 RALRWGRPRLAGLLCLAWVLVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPEEFDHYVHFS 201
         |: ||| | |: |||| | | ||| |||||||||| :||:||||:|: || |||| |
Db  140 HSLRWGRARYARRVAAVVWVLVLACQAPVLYFVTTSVRGTRITCHDTISARELFSHFVAYS 199

Qy  202 SAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRL--RSLRTIAVVLTVFAVCEV 259
         |||||| :|||||| | |: |||||||| :||:|| |||||||:|| || |:
Db  200 SVMLGLLFAVPFSVILVCYVLMARRLLKPAYGTTGGLPRAKRKSVRTIALVLAVFALCFL 259

Qy  260 PFHITRTIYYLARLLEADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLLTGDKYRRQLRQ 319
         |||||: :||| ||| || || |: |:|||||||||||||| ||: |:  ||:| |
Db  260 PFHVTRLLYYSFRSLDLSCHTLNAINAYKITRPLASANSCLDPVLYIFLAGQRLVRFARD 319

Qy  320 LCGGGKPQP----------RTAASSLALVSLPEDSSCRWAATPQDS 355
         :|:  | |          |:|| |       |          |
Db  320 AKPPTEPPPSPQARRKLGLHRPNRTVRKDL---SVSSDDSRRIESTPAGS 366

Qy  320 LCGGGKPQPRTAA 332
         |||||||||||
Db  318 AKPPPGPSPATPA 330
```

RESULT 3
A54946
P-2U nucleotide receptor - human
C;Species: Homo sapiens (man)
C;Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 17-Mar-1999
A;Accession: A54946
R;Parr, C.E.; Sullivan, D.M.; Paradiso, A.M.; Lazarowski, E.R.; Burch, L.H.; Olsen, J.C.
Proc. Natl. Acad. Sci. U.S.A. 91, 3275-3279, 1994
A;Title: Cloning and expression of a human P-2U nucleotide receptor, a target for cystic
R;Status: preliminary
A;Reference number: A54946; MUID:94211846
A;Accession: A54946
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-375 <PAR>
A;Cross-references: GB:U07225
A;Superfamily: ATP receptor P2u
C;Note: parts of this sequence were confirmed by protein sequencing
C;Genetics:
A;Gene: GDB:P2RY2; HP2U; P2U
A;Cross-references: GDB:362713; OMIM:600041
A;Map position: 11q13.5-11q14.1
A;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match   48.0%; Score 934; DB 2; Length 375;
Best Local Similarity 58.5%; Pred. No. 3.2e-75;
Matches 183; Conservative 41; Mismatches 85; Indels 4; Gaps 3;

```
Qy   22 EVELDCWFDEDKFPILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDKATATYMFHLALSD  81
         |: |||||||||||| ||||||||||||||||||||||||||:||||||:|
Db   20 ELGYKCRFNEDKFXVLLPVSYGVVCVLGICLNAVGLYIFLCRLKTWNASTTTMFHLAVSD  79
```

RESULT 4
JC4737
P2Y1 receptor G-protein coupled receptor-1 - human
N;Alternate names: P2Y1 purinergic receptor; P2Y1 purinoceptor
C;Species: Homo sapiens (man)
C;Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 24-Sep-1999
C;Accession: JC4737; JC4615
R;Janssens, R.; Communi, D.; Pirotton, S.; Samson, M.; Parmentier, M.; Boeynaems, J.M
Biochem. Biophys. Res. Commun. 221, 588-593, 1996
A;Title: Cloning and tissue distribution of the human P2Y1 receptor.
A;Reference number: JC4737; MUID:96205320
A;Accession: JC4737
A;Molecule type: DNA
A;Residues: 1-373 <JAN>
A;Cross-references: GB:S81950; NID:g1839438; PIDN:AABA70091.1; PID:g1839439
R;Ayyanathan, K.; Webbs, T.E.; Sandhu, A.K.; Athwal, R.S.; Barnard, E.A.; Kunapuli, S
Biochem. Biophys. Res. Commun. 218, 783-788, 1996
A;Title: Cloning and chromosomal localization of the human P2Y1 purinoceptor.
A;Reference number: JC4615; MUID:96158962
A;Accession: JC4615
A;Molecule type: mRNA
A;Residues: 1-373 <AYY>
A;Cross-references: GB:U42029; NID:g1147730; PIDN:AAA97872.1; PID:g1147731
A;Experimental source: erythro leukemia cells
C;Comment: This receptor belongs to a family of G protein-coupled receptors. It respo
C;Genetics:
A;Gene: P2Y1
A;Map position: 3
C;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane p
F;52-77/Domain: transmembrane #status predicted <TM1>
F;88-111/Domain: transmembrane #status predicted <TM2>
F;124-152/Domain: transmembrane #status predicted <TM3>
F;171-191/Domain: transmembrane #status predicted <TM4>
F;214-237/Domain: transmembrane #status predicted <TM5>
F;261-282/Domain: transmembrane #status predicted <TM6>
F;305-328/Domain: transmembrane #status predicted <TM7>
F;11,27,113,197/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;258,336/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status
F;330,339/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status pred
F;343/Binding site: phosphate (Ser) (covalent) (by protein kinase C and calmodulin-de

Query Match   33.0%; Score 641.5; DB 2; Length 373;
Best Local Similarity 40.0%; Pred. No. 2.3e-49;
Matches 140; Conservative 60; Mismatches 121; Indels 29; Gaps 9;

```
Qy   17 GPGSS----------------EVELDCWFDE-DFKFILLPVSYAVVFVLGLGLNAPTLWLFIFR  63
         ||||                 |||||||||       :|||||||||||||||||
Db   20 GPGSSWGNSTVASTAAVSSSFKCALTKTGFQFYLPAVYILVFIIGFLGNSVAIWMFVFH  79
```

Qy   64 LRPWDATATYMFHLALSDTLYVLSLPFTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSV 123
         |||||||| ||||:||:| | |:|:|| ||:|| |||| ||||||||||||||||||||
Db   80 MKPWSGISVYMFNLALADFLYVLTLPALIFYYFNKTDWIFGDAMCKLQRFIFHVNLYGSI 139

Qy   124 LFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLA--VWLVVAGCLVPNLFFVTTS-NKG 180
         ||||| ||||:|| | |:| | || | ||:||:| ||:|| || ||:|||| : |
Db   140 LFLTCISAHRYSGVVYPLKSL--GRLKKKNAICISVLVLWLIVVVAISPILFYSGTGVRKN 197

Qy   181 TTVLCHDTTRPEEFDHVYHFSSAVMGLLFGVPCLVTLVCYGLMARRL--YQPLPGSAQSSS 239
         |:|| :| ||::|: |:|| || || ||: : ||| |:|||: ||||||||||:
Db   198 KIITCYDTTSDEYLRSYFIYSMCTTVAMFCVPLVLILGCYGLIVRALIYKDLD---NSPL 254

Qy   240 RLRSLRTIAVVLLVFAVCFVPFHITRTIYYLARL--LEADCRVLNIVNVVKVTRPLLAS 296
         | |||: |||:| |:| |||||:|:| |:|: |||| |||:|||:|:|| ||
Db   255 RRKSIYLVIILVFAVSYIPFHVMKTMNLRARLDFQTPAMCAFNDRVYATYQVTRGLAS 314

Qy   297 ANSCLDPVLYLLTGDKYRRQL----ROLCGGGKPQPRTAASSIALVSLPE 342
         ||| || ||:||:|| |||: |::| || | | :: :| :| ||
Db   315 LNSCVDPILYFLAGDTFRRRLSRATRKASRSEANLQSKSEDMTLNILPE 364

RESULT 5
S54253
purinergic receptor - human
C;Species: Homo sapiens (man)
C;Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 24-Sep-1999
C;Accession: S54253
R;Leon, C.; Vial, C.; Cazenave, J.; Gachet, C.
A;submitted to the EMBL Data library, May 1995
A;Description: Cloning of a human putative P2Y receptor.
A;Reference number: S54253
A;Accession: S54253
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-372 <LEO>
A;Cross-references: EMBL:Z49205; NID:g798835; PIDN:CAA89066.1; PID:g798836
A;Genetics:
A;Gene: GDB:P2RY1
A;Cross-references: GDB:677125; OMIM:601167
A;Map position: 3pter-3qter
C;Superfamily: ATP receptor P2u

Query Match   32.4%; Score 630; DB 2; Length 372;
Best Local Similarity 39.7%; Pred. No. 2.4e-48;
Matches 139; Conservative 60; Mismatches 121; Indels 30; Gaps 10;

Qy   17 GPGSS------------------EVELDCWFDE-DFKFILLPVSYAVVFVLGLGLNAPTLWLFIFR 63
         ||||| :|:|||
Db   20 GPGSSWGNSTVASTAAVSSSSFKCALTKTGFQFYYLPAVYILVFIIGFLGNSVAIWMVFVH 79

Qy   64 LRPWDATATYMFHLALSDTLYVLSLPFTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSV 123
         |||||||| ||||:||:| | |:|:|| ||:|| |||| ||||||||||||||||||||
Db   80 MKPWSGISVYMFNLALADFLYVLTLPALIFYYFNKTDWIFGDAMCKLQRFIFHVNLYGSI 139

Qy   124 LFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLA--VWLVVAGCLVPNLFFVTTS-NKG 180
         ||||| ||||:|| | |:| | || | ||:||:| ||:|| || ||:|||| : |
Db   139 LFLTCISAHRYSGVVYPLKSL--GRLKKKNAICISVLVLWLIVVVAISPILFYSGTGVRKN 196

Qy   181 TTVLCHDTTRPEEFDHVYHFSSAVMGLLFGVPCLVTLVCYGLMARRL--YQPLPGSAQSSS 239
         |:|| :| ||::|: |:|| || || ||: : ||| |:|||: ||||||||||:
Db   181 KIITCYDTTSDEYLRSYFIYSMCTTVAMFCVPLVLILGCYGLIVRALIYKDLD---NSPL 253

Qy   240 RLRSLRTIAVVLLVFAVCFVPFHITRTIYYLARL--LEADCRVLNIVNVVKVTRPLLAS 296
         | |||: |||:| |:| |||||:|:| |:|: |||| |||:|||:|:|| ||
Db   254 RRKSIYLVIILVFAVSYIPFHVMKTMNLRARLDFQTPAMCAFNDRVYATYQVTRGLAS 313

Qy   297 ANSCLDPVLYLLTGDKYRRQL----ROLCGGGKPQPRTAASSIALVSLPE 342
         ||| || ||:||:|| |||: |::| || | | :: :| :| ||
Db   314 LNSCVDPILYFLAGDTFRRRLSRATRKASRSEANLQSKSEDMTLNILPE 363

RESULT   6

S33733
G protein-coupled receptor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C;Accession: S33733
R;Webb, T.E.; Simon, J.; Krishek, B.J.; Bateson, A.N.; Smart, T.G.; King, B.F.; Burns
FEBS Lett. 324, 219-225, 1993
A;Title: Cloning and functional expression of a brain G-protein-coupled ATP receptor.
A;Reference number: S33733; MUID:93285340
A;Accession: S33733
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-362 <WEB>
A;Cross-references: EMBL:X73268; NID:g395084; PIDN:CAA51716.1; PID:g395085
C;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match   32.3%; Score 628; DB 2; Length 362;
Best Local Similarity 39.4%; Pred. No. 3.5e-48;
Matches 137; Conservative 62; Mismatches 133; Indels 16; Gaps 7;

Qy   4 TESLLLRSLGLSPGSSSEVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFR 63
         | ||| | | :|||||| |||||||||:||:| | |:|:||
Db   13 TQPELLAG-GWAAGNATTKCSLT---KTGFQFYYLPTVILVFIITGFLGNSVAIWMVFVH 68

Qy   64 LRPWDATATYMFHLALSDTLYVLSLPFTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSV 123
         |||||||| ||||:||:| | |:|:|| ||:|| |||| ||||:|||||||||||||
Db   69 MRPWSGISVYMFNLALADFLYVLTLPALIFYYFNKTDWIFGDVMCKLQRFIFHVNLYGTI 128

Qy   124 LFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVWLVVAGCLVPNLFFVTTS-NKGTI 182
         |||||||||||:| | |:| | ||| | | |: || || : |:|:|||| : |:|
Db   129 LFLTCISVHRYTGVVHPLKSLGRLKKKNAVVSSLVWALVVAVIRAPILFYSGTGVRRNKT 188

Qy   183 VLCHDTTRPEEFDHVYHFSSAVMGLLFGVPCLVTLVCYGLMARRL--YQPLPGSAQSSSRL 241
         :|| :| ||::|: |:|| | || ||: : ||| |:|||: |||||||||:|
Db   189 ITCYDTTADEYLRSYFVYSMCTTVPMFCIPFIVILGCYGLIVKALIYKDLD---NSPLRR 245

Qy   242 RSLRTIAVVLLVFAVCFVPFHITRTIYYLARL--LEADCRVLNIVNVVKVTRPLAS 298
         | |: |||:| |:| |||||:|:| |:|: |||| |||:|||:|:||
Db   246 KSIYLVIILVLTVFAVSYLPFHVMKTLNLRARLDFQTPQMCAFNDKVYATYQVTRGLASLN 305

Qy   299 SCLDPVLYLLTGDKYRRQL----ROLCGGGKPQPRTAASSIALVSLPE 342
         ||| || ||:||:|| |||: |::| || | | :: :| :| ||
Db   306 SCVDPILYFLAGDTFRRRLSRATRKASRSEPNVQSKSEEMTLNILTE 353

RESULT   7
JC4162
P2Y receptor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 12-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 24-Sep-1999
C;Accession: JC4162
R;Henderson, D.J.; Elliot, D.G.; Smith, G.M.; Webb, T.E.; Dainty, I.A.
Biochem. Biophys. Res. Commun. 212, 648-656, 1995
A;Title: Cloning and characterisation of a bovine P2Y receptor.
A;Reference number: JC4162; MUID:95352058
A;Accession: JC4162
A;Molecule type: mRNA
A;Residues: 1-373 <HEN>
A;Cross-references: EMBL:X87628; NID:g1032484; PIDN:CAA60958.1; PID:g1032485
A;Experimental source: aortic endothelial cell
A;Genetics:
A;Gene: bovp2y
C;Superfamily: ATP receptor P2u
C;Superfamily: bovp2y
C;Keywords: glycoprotein; phosphoprotein; receptor; transmembrane protein
F;52-77/Domain: transmembrane #status predicted <TM1>
F;88-111/Domain: transmembrane #status predicted <TM2>
F;124-150/Domain: transmembrane #status predicted <TM3>
F;171-191/Domain: transmembrane #status predicted <TM4>
F;214-237/Domain: transmembrane #status predicted <TM5>
F;261-282/Domain: transmembrane #status predicted <TM6>
F;305-328/Domain: transmembrane #status predicted <TM7>

F;11,27,113,197/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;258/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted

```
Query Match                31.9%;     Score 621;  DB 2;    Length 373;
Best Local Similarity     43.4%;  Pred. No. 1.5e-47;
Matches 126;  Conservative  53;  Mismatches 103;  Indels   8;  Gaps  4;

Qy    33 FKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLI 92
            |:|  |  |  ||:|||  |:||:| :|||||||||||||||:||||||||||||||||
Db    49 FQFYLPAVYLVFIIGFLGNSVAIWMFVFHMKPWSGISVTMFNLALADFLYVILTLPALI 108

Qy    93 YYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLA 152
            :|| |||||||||||:||||||||||||||||||||| :||||: :|  :|
Db   109 FYYFNKIDWIFGDAMCKLQRFIFHVNLYGSILFLTCISAHRYSGVVYPLKSLGRLKKKNA 168

Qy   153 GLICLAVWLVVAGCLVPNLFEVTTS-NKGTTVLCHDTTRPEEFDHYVHFSSAVMGLLFGV 211
            |||  |  || |  ||||||:||: : ::| |   ::|  ::||||:||||||||||||
Db   169 VYISLVLWLIVVVGISPILFYSGTGIRKNKIITCYDTTSDEYLRSYFIYSMCTTVAMFCV 228

Qy   212 PCLVTLVCYGLMARRL-YQPLPGSAQSSSRLRSLRTIAVLTVFAVCFVPFHITRTIIYL 270
            ||:||||||||||||  ::||:||||||||||||||||||||:||||||||:||||:|
Db   229 PLVLILGCYGLIVRALIYKDLD---NSPLRRKSIVLIVIVLTVFAVSYIPFHVMKTMNLR 285

Qy   271 ARL---LEADCRVLNIVNVVKVTRPLASANSCLDPVLLLTGDKYRRQL 317
            |||        ||||:::::|: ||: :|| :|:|||||||||||||
Db   286 ARLDFQTPEMCAFNDRVYATYQVTRGLASLNSCVDPILYFLAGDTFRRRL 335
```

```
RESULT    8
I55450
a novel G protein-coupled P2 receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Jul-1996  #sequence_revision 02-Jul-1996  #text_change 24-Sep-1999
C;Accession: I55450
R;Chang, K.; Hanaoka, K.; Kumada, M.; Takuwa, Y.
J. Biol. Chem. 270, 26152-26158, 1995
A;Title: Molecular Cloning and functional analysis of a novel P2 nucleotide receptor.
A;Reference number: I55450; MUID:96064682
A;Accession: I55450
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-328  <RES>
A;Cross-references: GB:D63665; NID:g1066007; PIDN:BAA09816.1; PID:g1066008
C;Superfamily: G protein-coupled receptor
```

```
Query Match                30.2%;     Score 588;  DB 2;    Length 328;
Best Local Similarity     42.8%;  Pred. No. 1.1e-44;
Matches 136;  Conservative  37;  Mismatches 133;  Indels  16;  Gaps  6;

Qy     5 ESSLLRSLGLSPGPGSSEVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRL 64
         :: :||:|:|   |||:  ::||||::||:||||||||||||||||||||||||||||||
Db     4 DNGTIOAPGLPP------TTCVIREDFKRLLPVYSVVLVVGLPLNVCVIAQICASR 55

Qy    65 RPWDATATYMFHLALSDTLYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVL 124
         ::|:|:|||||||||||||:||||||||:  ||||||||||||||||||||||||||||
Db    56 RTLTRSAVYTLNLALADLLYACSLPLLIYNYARGDHWPFGDLACRLVRFLFYANLHGSIL 115

Qy   125 FLTCISVHRYLGICHPLRALRW--GRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGT 181
         |||||||||||||||||||||    ||||| :  |||:|||||||||||||||||||
Db   116 FLTCISFQRYLGICHPLAP--WHKRGGRRAAWVCGVWLVVTAQCLPIAVFAATGIQRN 173

Qy   182 TVLCHDTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRL-YQPLP-GSAQSSS 239
         |:|   |:|  ::||||:||||||||||||||:|||||||||||| |||||  ||||||
Db   174 RTVCVDLSPPILSTRYLPYGMALTVIGFLLPFTALLACTCRMARRLCRODGPAGPVAQER 233

Qy   240 RLRSLRTIAVLTVFAVCFVPFHITRTIIYLARLLE-ADCRVLNIVNVVKVTRPLASAN 298
         ||||||||||||||||  ||||||||||||||||  |||||||:|||||||||||||
Db   234 RSKAARMAVVAAAFVISFLPFHITKTAYLAVRSTPGVSCPVLETFAAAYKGTRPFASAN 293

Qy   299 SCLDPVLLLTGDKYRRQLRQL 320
         |||||:||||||||||||||
Db                PILFYFTQKKFRRPHEL 313
```

```
RESULT    9
G02514
P2 purinoceptor - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996  #sequence_revision 06-Jun-1997  #text_change 24-Sep-1999
C;Accession: G02514
R;Hammet, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J.
submitted to the EMBL Data Library, March 1996
A;Reference number: H01373
A;Accession: G02514
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-326  <HAM>
A;Cross-references: EMBL:U52464; NID:g1407632; PIDN:AAB03572.1; PID:g1407633
C;Genetics:
A;Gene: P2Y6
C;Superfamily: ATP receptor P2u
```

```
Query Match                30.1%;     Score 586;  DB 2;    Length 326;
Best Local Similarity     42.5%;  Pred. No. 1.6e-44;
Matches 135;  Conservative  37;  Mismatches 128;  Indels  18;  Gaps  6;

Qy    10 RSLGLSPGPGSSEVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDA 69
         ::|||    |||:  ::||||::||:||||||||||||||||||||||||||||||
Db     7 QALGLPP------TTCVIRENFKQLLLPPVYSAVLAAGLPLNICVITOICTSRRALTR 58

Qy    70 TATYMFHLALSDTLYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCI 129
         |:|||||||||||||:|||||||  ||||||||||||||||||||||||||||||||||
Db    59 TAVYTLNLALADLLYACSLPLLIYNYAQGDHWPFGDEFACRLVRFLFYANLHGSILFLTCI 118

Qy   130 SVHRYLGICHPLRALRW---GRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCH 186
         |:|||||||||||||    ||||| :|  |||:|||||||||||||||||||||||:
Db   119 SFQRYLGICHPLAP--WHKRGGRRAAWLVCVAVWLAVTTQCLPTAIFAATGIQRNRTVCI 176

Qy   187 DTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQ---PLPGSAQSSSRLRS 243
         |:|  ::||||:||||||||||||||:||||||||||||| |     ||||||||||||
Db   177 DLSPPALATHYMPYGMALTVIGFLLPFAALLACYCLLACRLCRODGPAEPVAQ--ERGKA 235

Qy   244 LRTIAVLTVFAVCFVPFHITRTIIYLARLLE-ADCRVLNIVNVVKVTRPLASANSCLD 302
         |||||||||||||  |||||||||||||||  |||||||:|||||||||||||||
Db   236 ARMAVVVAAAFAISFLPFHIRTKAYLAVRSTPGVPCTVLEAFAAAYKGTRPFASANSVLD 295

Qy   303 PVLLLTGDKYRRQLRQL 320
         |:|:|  ||||||||||
Db   296 PILFYFTQKKFRRPHEL 313
```

```
RESULT   10
JC4800
P2Y6 receptor - human
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1995  #sequence_revision 16-Aug-1996  #text_change 04-Sep-1998
C;Accession: JC4800
R;Communi, D.; Parmentier, M.; Boeynaems, J.M.
Biochem. Biophys. Res. Commun. 222, 303-308, 1996
A;Title: Cloning, functional expression and tissue distribution of the human P2Y6 rec
A;Reference number: JC4800; MUID:96222498
A;Accession: JC4800
A;Molecule type: mRNA
A;Residues: 1-328  <COM>
A;Cross-references: EMBL:X97058
A;Experimental source: placenta
C;Superfamily: ATP receptor P2u
C;Keywords: glycoprotein; placenta; receptor; transmembrane protein
F;26-52/Domain: transmembrane #status predicted <TMM1>
F;63-86/Domain: transmembrane #status predicted <TMM2>
F;104-122/Domain: transmembrane #status predicted <TMM3>
F;143-167/Domain: transmembrane #status predicted <TMM4>
```

F;193-216/Domain: transmembrane #status predicted <TMM5>
F;241-264/Domain: transmembrane #status predicted <TMM6>
F;283-305/Domain: transmembrane #status predicted <TMM7>
F;5,173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   30.1%; Score 586; DB 2; Length 328;
Best Local Similarity   42.5%; Pred. No. 1.6e-44;
Matches 135; Conservative 37; Mismatches 128; Indels 18; Gaps 6;

Qy   10 RSLGLSPGSSEVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDA 69
                :||:|:| ::|::|||||||||||||||||||||||||:|
Db   9 QALGLPP-------TTCVYRENFKQLLLPPVYISAVLAAGLPLNICVITQICTSRRALTR 61

Qy   70 TATYMFHLAI.SDTILVLSLPFLIIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCI 129
                :||||:||| ::|||||:|||||||||||||||||| ::||||||||:||:||||||
Db   61 TAVYILNLALADLLTYACSLPLLIYYANAQGDHWPFGDEACRLVRFLFYANLHGSILFLTCI 120

Qy   130 SVHRYLGICHPLRALRW---GRPRLAGLLCLAVWLVVAGCLVPNLFVTTSNKGTVTVLCH 186
                :||:|||||||||||| ||| :|||::|||| |||::|::|||||||||||:|||:
Db   121 SFQRYLGICHPLAP--WHKRGGRRAAWLVCVAVWLAVTTQCLPTAIFAAATGIQRNRTVCY 178

Qy   187 DTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQ---PLPGSAQSSSRLRS 243
                 |||||||||||||||||| |||||||||||:|||||||||| : ||:|||::|||:
Db   179 DLSPPALATHYMPYGMALTVIGFLLPPAALLACYCLLACRLCRQDGPAEPVAQ-ERRGKA 237

Qy   244 LRTIAVVLTVFAVCVPFHITRTIYYLARLLE-ADCRVLNIVNVVYKVTRPLASANSCLLD 302
                ||||||||:|||||||||||||||||||||| |||||||||||:||:|||||||:|:||
Db   238 ARMAVVVAAAFAISPLPFHITKTAVLAVRSTPGVPCTVLEAFAAAYKGTRPFASANSVLD 297

Qy   303 PVLILLTGDKYRRQLRQL 320
                ||:|:| :||:|:|
Db   298 PILFYFTQKKFRRPHEL 315

RESULT 11
I50241

G protein-coupled receptor 6H1 - chicken
N;Alternate names: purinoceptor 6H1
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
R;Kaplan, M.H.; Smith, D.I.; Sundick, R.S.
J; Immunol. 151, 628-636, 1993
A;Title: Identification of a G protein coupled receptor induced in activated T cells.
A;Reference number: I50241; MUID:93329058
A;Accession: I50241
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-308 <KAP>
A;Cross-references: GB:L06109; NID:g304383; PIDN:AAB06587.1; PID:g304384
R;Webb, T.E.; Kaplan, M.G.; Barnard, E.A.
Biochem. Biophys. Res. Commun 219, 105-110, 1996
A;Title: Identification of 6H1 as a P2Y purinoceptor: P2Y5.
A;Reference number: JC4618; MUID:96190677
A;Accession: JC4618
A;Molecule type: mRNA
A;Residues: 1-308 <WEB>
A;Cross-references: DDBJ:D49712
A;Experimental source: T-cells
C;Comment: This receptor plays a role in T-cell activation.
C;Genetics:
A;Gene: P2Y5
C;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor; transmembrane protein
F;15-40/Domain: transmembrane #status predicted <TM1>
F;51-74/Domain: transmembrane #status predicted <TM2>
F;89-109/Domain: transmembrane #status predicted <TM3>
F;133-153/Domain: transmembrane #status predicted <TM4>
F;177-201/Domain: transmembrane #status predicted <TM5>
F;227-248/Domain: transmembrane #status predicted <TM6>
F;269-292/Domain: transmembrane #status predicted <TM7>

Query Match   25.8%; Score 501.5; DB 2; Length 308;
Best Local Similarity   36.9%; Pred. No. 4.5e-37;
Matches 108; Conservative 56; Mismatches 124; Indels 5; Gaps 4;

Qy   23 VELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDT 82
                :| :| |:::| ||||||||||||||||||||||||||||:|||||| :|| ::||:|
Db   2 VSSNCSTEDSFKYTLYGCVFSMVFVLGLIANCVAIYIFTFLKVRNETTYTYMLNIAISDL 61

Qy   83 LYVLSLPFLIIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLR 142
                |:|:|||||||||:| |||| ::||||||||||||||||||||||||:||:||||||||
Db   62 LFVTLPFRIYIFVVRN-WPFGDVLCKISVSVTLFYTNMYGSILFLTCISVDRFLAIVHPER 120

Qy   143 ALRWGRPRLAGLLCLAVWLVVAGCLVPNLFVTTSNKGTVTVLCH 200
                ||||::|:|||::|||| ||||: |:|:|||||||||||:||::||:
Db   121 SKYLRTKRNARIVCVAVWTVLAGSTPASFFQSTNRQNNTEQRRCFENFPEESTWKTILSR 180

Qy   201 SSAVMGLL-FGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFV 259
                :||:|||| |||||||||:||||||||||| |||:|||::||||:|||||||||||||||
Db   181 IVIFIEIVGFFIPLLLLNVTCSMWLRTLNKPLILSRNKLSKKKVLKMIFVHLLVIFCFCFV 240

Qy   260 PFHITRTIYYLARLLE-ADCRVLNIVNVVYKVTRPLASANSCLLDPVLLLLTGD 311
                ||||||||||||||||| |||||||||||||||:||||||||:|:||:|:| |
Db   241 PYNTLLIYSLMRIQTWINCSVTAVRFMYPVTLCIAVSNCCFPDIVIVFTSD 293

RESULT 12
T09508

intron 17 purinergic receptor P2Y5 - human
N;Alternate names: G-protein coupled receptor
A;Accession: T09508
A;Reference number: Z16705
A;Status: preliminary; G-protein coupled receptor
A;Species: Homo sapiens (man)
A;Molecule type: mRNA
A;Residues: 1-344 <BOH>
A;Cross-references: EMBL:AF000546; NID:g2232068; PID:g2232069
C;Accession: T09508
R;Bohm, S.K.; Trumpp, A.; Khitin, L.M.; Kong, W.; Payan, D.G.; Bunnett, N.W.
submitted to the EMBL Data Library, April 1997
A;Description: The human purinergic receptor P2Y5 is encoded in intron 17 of the reti
A;Genetics:
A;Map position: 13
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match   24.6%; Score 479.5; DB 2; Length 344;
Best Local Similarity   33.4%; Pred. No. 4.4e-35;
Matches 99; Conservative 66; Mismatches 126; Indels 5; Gaps 4;

Qy   27 CWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVL 86
                |:||:| :| ||||||||||:||| |||||||||||:||:|:| :|| :||:||:|
Db   9 CFYNDSFKYTLYGCMFSMVFVLGLVSNCVAIYIFICVLKVRNETTYTYMINLAMSDLLFVF 68

Qy   87 SLPTLIIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRW 146
                ||| :||:| |||| :||||||||||||||||||||||||||: |:|:||||||||||||
Db   9 TLPFRIYFTTRN-WPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVPFKSKTL 127

Qy   147 GRPRLAGLLCLAVWLVVAGCLVPNLFVTTSNKGTTV--LCHDTTRPEEFDHYVHFSSAV 204
                ||||||||||||| ||| :|:|:|:||||||::| : |||||||||||||||::|||||
Db   128 RTKRNAKIVCTGVWLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIF 187

Qy   205 MGLL-FGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVCVPFHI 263
                |||| |||||||||:||||||||||| |||:|||::||||:|:|||||||||:| ::|
Db   188 IEIVGFFIPLILNVTCSSMVLKTILTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNI 247

Qy   264 TRIIYYLARLLE-ADCRVLNIVNVVYKVTRPLASANSCLLDPVLYLLIGDKYRRQLR 318
                |:|| | :|:: | |||||||||||||:|||:|:|:|: ||||:|:| |::
Db   248 NLIIYSLVRTQTFVNCSVVAAVRTMYPITLCIAVSNCCFPDIVIYYFTSDTIQNSIK 303

RESULT 13

JC5549
heptahelical P2Y5-like receptor - human
C;Species: Homo sapiens (man)
C;Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 24-Sep-1999
C;Accession: JC5549
R;Janssens, R.; Boeynaems, J.M.; Godart, M.; Communi, D.
Biochem. Biophys. Res. Commun. 236, 106-112, 1997
A;Title: Cloning of a human heptahelical receptor closely related to the P2Y5 receptor.
A;Reference number: JC5549; MUID:97366605
A;Accession: JC5549
A;Molecule type: DNA
A;Residues: 1-370 <JAN>
A;Cross-references: DDBJ:AF005419; NID:g2240034; PIDN:AAB66322.1; PID:g2240035
C;Superfamily: ATP receptor P2u

Query Match   23.2%;   Score 450.5; DB 2;   Length 370;
Best Local Similarity   33.4%;   Pred. No. 1.7e-32;
Matches 102; Conservative 67; Mismatches 131; Indels 5; Gaps 5;

Qy   14 LSPGPGGSSEVELDCWFDEDFKILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATY 73
          | || |: | | |:| |||||:|||||:||||||::|||: | | : | | |
Db   18 LRPRLGNATANNTCIVDDSFKYRLNGAVVSVVFILGLITNVSLFVFCFRMKRSETAIF 77

Qy   74 MFHLALSDTLYVLSLPFLIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHR 133
          : :|:|| | || | |||| : ||| | | |||||||:|||||||||||||||||
Db   78 ITNLAVSDLLFVCTLPFKI-FYNFNRHWPFGDFLCKISGTAFLTNIGSMLFLTCISVDR 136

Qy   134 YLGICHPLRALRWGRPRLAGLLCLAVW--LVVAGCLVPNLFVFVTTSNKGTTVLCHDTRPE 192
           ||||||||||| | | | :| ||| | || |: | |::| ||||||| |
Db   137 FLAIVFPFRSIRTRTRNSAIVCAGVWILVLSGGISASLFSTTNVNNATTT-CFEGLSKR 195

Qy   193 EFDHYVHFSSAVMGLL-FGVPCLVTLVCYGLMARRLYQPLPGSSAQSSSRLRSLRTIAVVL 251
          | |:| ||| |:| |||||:||||||||||||||||||||||||||||||| | ||
Db   196 VWKTYLSKITIFIEVVGFIIPLLLNVSCSSVLFTLRKPATLSQIGTNKKKVLKMITVHM 255

Qy   252 TVFAVCFVPFHITRTIYYLIARLLE-ADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTG 310
          | | | || | : ||::|||:| ||||||||||||||||||||||||||||||||||
Db   256 AVFVCVCFVPYNSVLFLYALVRSQAITNCFLEREFAKIMYPITLCLATLNCCFDPIYYFTL 315

Qy   311 DKYRR 315
          : ::|
Db   316 ESFQK 320

RESULT 14
I48705
proteinase activated receptor 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Jul-1996
C;Accession: I48705
R;Nystedt, S.; Larsson, A.K.; Aberg, H.; Sundelin, J.
J. Biol. Chem. 270, 5950-5955, 1995
A;Title: The mouse proteinase-activated receptor-2 cDNA and gene. Molecular cloning and
A;Reference number: I48705; MUID:95197620
A;Accession: I48705
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-399 <RES>
A;Cross-references: EMBL:Z48043; NID:g663020; PID:g663021

Query Match   21.7%;   Score 423; DB 2;   Length 399;
Best Local Similarity   31.8%;   Pred. No. 5e-30;
Matches 107; Conservative 55; Mismatches 141; Indels 34; Gaps 10;

Qy   36 ILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLIYY 95
          :||||:| ||||| ||||||||||||:|| |||:|| | || ||| :|:|||:| |||
Db   78 VFLPVVIVIFVIGLPSNGMALWIFLFRTKKKHPAVIMANLALADLLSVIWFPLKISYH 137

Qy   96 AAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPR---L 151
          |||:|||||| : | |||| || :|:|||| ||| ||||||||||||||:| : |
Db   138 LHGNNWVYGEALCKVLIGFFYGNMYCSILFMTCLSVQRYWVIVNPM-----GHPRKKANI 192

RESULT 15
I51667
thrombin receptor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C;Accession: I51667
R;Gerszten, R.E.; Chen, J.; Ishii, M.; Ishii, K.; Nanevicz, T.; Turck, C.W.; Vu, T.H.
Nature 368, 648-651, 1994
A;Title: Thrombin receptor's specificity for agonist peptide is determined by its ext
A;Reference number: I51667; MUID:94195429
A;Accession: I51667
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-420 <GER>
A;Cross-references: EMBL:U09632; NID:g495197; PID:g495198

Query Match   21.1%;   Score 411.5; DB 2;   Length 420;
Best Local Similarity   29.7%;   Pred. No. 5.5e-29;
Matches 98; Conservative 61; Mismatches 144; Indels 27; Gaps 7;

Qy   34 KFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLIIY 93
          | ||||:| | ||:|| |||||||||:||:||:|||:||:| |:|||| |||:|||:|||
Db   102 KFV--PSLYTVVFVIGLPLNLLAIIIFLFKMVRKPAVVMLNLAIADVFFVSVLPFKIA 159

Qy   94 YYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAG 153
          | ::| :|||||| : | ||| || |||:||| || :|||||||||||| | | |
Db   160 YHLSGNDWLFGPGMCRIVTAIFYCNMYCSVLLIASISVDRFLAVVYPMHSLSWRTMSRAY 219

Qy   154 LLCLAVWLVVAGCLVPNLFVFVTSN--KGTTVLCHDTRPEEF-DHYVHFSSAVMGLLFLG 210
          :||| ||:| || |||| : || || ||||||||: | |||:| ||| || |
Db   220 MACSFIWLISIASTPLLVTEQTOKIPRLDITTCHDVLDLKDLKDFYIYFSSPCLLFFF 279

Qy   211 VPCLVTLVCYGLMARRLYQPLPGSSAQSSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYYL 270
          ||||:||||||||||||||||||||||||||||||||| || | :|||||||:|:| |:||
Db   280 VPFITTICYIGIIRSLSS--SSIENSCKKTRALFLAVVVLCVFIICFGPTNVLFLTHYL 337

Qy   271 ARLLEADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTGDKYRRQLRQ-LCGGGKPQPR 329
          |: |||||||||||||||||||||||||||||| : || || :| |:| | | | |
Db   338 QEANE------FLYFAYILSACVGSVSCCLDPLIYYASSQCORYLSLLCCRKVSBPG 390

Qy   330 TAASSLALVSLPEDSSCRWAATPQDSSSGST 359
          |||| ||||||||| | | | ||
Db   391 SSTGQLMSTAMKND-----------------NCST 408

Search completed: December 29, 1999, 17:46:05
Job time: 2592 sec

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: December 29, 1999, 19:43:41 ; Search time 26.5 Seconds
            (without alignments)
            400.618 Million cell updates/sec

Title: US-09-077-173A-2
Perfect score: 1946
Sequence: 1 MASTESSLLRSLGLSPGCGS...........CRWAATPQDSSCSTPRADRF 365
Scoring table: BLOSUM62
Searched: 80000 seqs, 29085965 residues
Database :   SwissProt_38:*

Word size :   0

Number of hits that pass the threshold :   80000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1940	99.7	365	1	P2Y4_HUMAN	P51582 homo sapien
2	1007.5	51.8	537	1	P2Y8_XENLA	P79928 xenopus lae
3	965	49.6	377	1	P2UR_HUMAN	P41231 homo sapien
4	962.5	49.5	373	1	P2UR_MOUSE	P35383 mus musculu
5	950	48.8	374	1	P2UR_RAT	P41232 rattus norv
6	641.5	33.0	373	1	P2YR_HUMAN	P47900 homo sapien
7	631	32.4	362	1	P2YR_MELGA	P49652 meleagris g
8	628	32.3	362	1	P2YR_CHICK	P34996 gallus gall
9	621	31.9	373	1	P2YR_BOVIN	P48042 bos taurus
10	616	31.7	373	1	P2YR_RAT	P49651 rattus norv
11	614	31.6	373	1	P2YR_MOUSE	P49650 mus musculu
12	603	31.0	328	1	P2Y3_CHICK	P98907 gallus gall
13	588	30.2	328	1	P2Y6_RAT	Q63371 rattus norv
14	586	30.1	328	1	P2Y6_HUMAN	Q15077 homo sapien
15	510	26.2	339	1	GPRH_HUMAN	Q13304 homo sapien
16	501.5	25.8	308	1	P2Y5_CHICK	P32250 gallus gall
17	479.5	24.6	344	1	P2Y5_HUMAN	P43657 homo sapien
18	449.5	23.1	370	1	P2Y9_HUMAN	Q99677 homo sapien
19	426	21.9	397	1	P2R2_RAT	Q63645 rattus norv
20	423	21.7	399	1	P2R2_MOUSE	P55086 rattus norv
21	411.5	21.1	420	1	PAR2_XENLA	P47749 xenopus lae
22	403	20.7	397	1	THRR_XENLA	P55085 xenopus lae
23	397.5	20.4	425	1	PAR2_HUMAN	P55088 homo sapien
24	393.5	20.2	425	1	THRR_HUMAN	P25116 homo sapien
25	393	20.2	425	1	THRR_PAPHA	P56488 papio hamad
26	388	19.9	430	1	THRR_MOUSE	P30558 mus musculu
27	387.5	19.9	432	1	THRR_RAT	P26824 rattus norv
28	385	19.8	428	1	THRR_CRILO	O00991 cricetulus
29	384	19.7	372	1	OPRD_RAT	P33533 rattus norv
30	383	19.7	372	1	OPRD_HUMAN	P41143 homo sapien
31	382.5	19.7	361	1	EB12_HUMAN	P32249 homo sapien
32	380.5	19.6	362	1	AG2R_XENLA	P32303 xenopus lae
33	380	19.5	401	1	OPRM_PIG	Q95247 sus scrofa
34	378.5	19.5	401	1	OPRM_HUMAN	P79350 homo sapien
35	377	19.4	400	1	OPRM_BOVIN	P35372 bos taurus
36	376.5	19.3	398	1	OPRD_MOUSE	P33300 mus musculu
37	375.5	19.3	374	1	OPRM_RAT	P33535 rattus norv
38	375	19.3	368	1	PAR3_HUMAN	O00254 homo sapien
39	371.5	19.1	398	1	CCR3_HUMAN / OPRM_MOUSE	Q49682 / P42866 homo sapien / mus musculu
40	371.5	19.1	342	1	PAFR_HUMAN	P25105 homo sapien
41	369.5	19.0	359	1	AG2R_MOUSE	P29754 mus musculu
42	367	18.9	357	1	GC96_HUMAN	P51686 homo sapien
43	365	18.8	363	1	SSR5_HUMAN	P35346 homo sapien
44	364.5	18.7	359	1	AG2R_RAT	P25095 rattus norv
45	363	18.7	359	1	AG2S_RAT	P29089 rattus norv

ALIGNMENTS

RESULT 1
P2Y4_HUMAN

ID   P2Y4_HUMAN   STANDARD;   PRT;   365 AA.
AC   P51582;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   URIDINE NUCLEOTIDE RECEPTOR (UNR) (P2P) (P2Y4).
GN   P2RY4 OR NRU.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96125054.
RA   NGUYEN T., ERB L., WEISMAN G.A., MARCHESE A., HENG H.H.Q.,
RA   GARRAD R.C., GEORGE S.R., TURNER J.T., O'DOWD B.F.;
RT   "Cloning, expression, and chromosomal localization of the human
RT   uridine nucleotide receptor gene.";
RL   J. Biol. Chem. 270:30845-30848(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 96197801.
RA   STAM N., KLOMP J., VAN DER HEUVEL M., OLIVE W.;
RT   "Molecular cloning and characterization of a novel orphan receptor
RT   (P2P) expressed in human pancreas that shows high structural homology
RT   to the P2U purinoceptor.";
RL   FEBS Lett. 384:260-264(1996).
RN   [2]
RA   COMMUNI D., PIROTTON S., PARMENTIER M., BOEYNAEMS J.-M.;
RT   "Cloning and functional expression of a human uridine nucleotide
RT   receptor.";
RL   J. Biol. Chem. 270:30849-30852(1995).
-!- FUNCTION: RECEPTOR FOR UTP AND UDP, NOT ACTIVATED BY ATP OR ADP.
     SEEMS TO MEDIATE ITS ACTION VIA ACTIVATION OF A
     PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- TISSUE SPECIFICITY: PANCREAS.
-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

DR   EMBL; X91852; CAA62963.1; -.
DR   EMBL; U40223; AAC50347.1; -.
DR   EMBL; X96597; CAA65415.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1952; -.
DR   GCRDB; GCR_2065; -.
DR   GCRDB; GCR_2085; -.
MIM; 300038; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.

```
FT   DOMAIN        1     34   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     35     61   1 (POTENTIAL).
FT   DOMAIN       62     72   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     73     95   2 (POTENTIAL).
FT   DOMAIN       96    112   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    113    131   3 (POTENTIAL).
FT   DOMAIN      132    154   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    155    174   4 (POTENTIAL).
FT   DOMAIN      175    196   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    197    222   5 (POTENTIAL).
FT   DOMAIN      223    246   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    247    269   6 (POTENTIAL).
FT   DOMAIN      270    287   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    288    309   7 (POTENTIAL).
FT   DOMAIN      310    365   CYTOPLASMIC (POTENTIAL).
FT   DISULFID    108    185   BY SIMILARITY.
FT   CONFLICT     86     86   L -> V (IN REF. 2).
FT   CONFLICT    178    178   N -> T (IN REF. 2).
FT   CONFLICT    234    234   S -> A (IN REF. 2).
FT   SEQUENCE    365 AA;  40963 MW;  AEE553A7 CRC32;
SQ
```

```
Query Match             99.7%;  Score 1940;  DB 1;  Length 365;
Best Local Similarity   100.0%;  Pred. No. 2.5e-129;
Matches  364;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MASTESSLLRSLGLSPGPGSSEVELDCWFDEDEPKFILLPVSIAVVFVLGLGLNAPTLWLF 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MASTESSLLRSLGLSPGPGSSEVELDCWFDEDEFKFILLPVSTAVVFVLGLGLNAPTLWLF 60

Qy   61 IFRLRPWDATATYMFHLALSDTLYVLLSLPTLIYYYAARHNWPFGTEICKFVRFLFYWNLY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 IFRLRPWDATATYMFHLALSDTLYVLLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLY 120

Qy  121 CSVLFLTCISVHRYLGICHPLRALRWGPRPLAGLLCLAVWLVVAGCLVPNLFFVTTSNKG 180
         |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  121 CSVLFLTCISVHRYLGICHPLRLRWGPRPLAGLLCLAVWLVVAGCLVPNLFFVTTSNKG 180

Qy  181 TTVLCHDTTRPEEFDHVVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSR 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTVLCHDTTRPEEFDHVVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSR 240

Qy  241 LRSLRTIAVLTVFAVCFVPFHITRTIYYLARLLEADCRVLNIVNVYKVTRPLASANSC 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LRSLRTIAVLTVFAVCFVPFHITRTIYYLARLLEADCRVLNIVNVYKVTRPLASANSC 300

Qy  301 LDPVLYLLTGDKYRRQLROLCGGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCSTP 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 LDPVLYLLTGDKYRRQLROLCGGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCSTP 360

Qy  361 RADR 364
         ||||
Db  361 RADR 364
```

```
RESULT 2
P2Y8_XENLA
ID   P2Y8_XENLA     STANDARD;      PRT;   537 AA.
AC   P79928;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   P2Y PURINOCEPTOR 8 (P2Y8).
GN   P2Y8.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=NEURAL PLATE;
RX   MEDLINE; 97284734.
RA   BOGDANOV Y.D., DALE L., KING B.F., WHITTOCK N., BURNSTOCK G.;
RT   "Early expression of a novel nucleotide receptor in the neural plate
RT   of Xenopus embryos.";
RL   J. Biol. Chem. 272:12583-12590(1997).
```

-!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ATP, UTP, CTP, GTP AND ITP.
    THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH
    ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
    MAY PLAY A KEY ROLE IN THE EARLY DEVELOPMENT OF NEURAL TISSUE.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

```
DR   EMBL; X99953; CAA68213.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1419; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     49   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     50     70   1 (POTENTIAL).
FT   DOMAIN       71     79   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     80    100   2 (POTENTIAL).
FT   DOMAIN      101    118   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    119    139   3 (POTENTIAL).
FT   DOMAIN      140    161   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    162    182   4 (POTENTIAL).
FT   DOMAIN      183    210   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    211    231   5 (POTENTIAL).
FT   DOMAIN      232    254   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    255    275   6 (POTENTIAL).
FT   DOMAIN      276     ?    7 (POTENTIAL).
FT   TRANSMEM     ?      ?    EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       ?      ?    7 (POTENTIAL).
FT   DISULFID    116    193   BY SIMILARITY.
FT   CYTOPLASMIC (POTENTIAL).
FT   BY SIMILARITY.
FT   POTENTIAL.
FT   POTENTIAL.
FT   CARBOHYD     26     26   POTENTIAL.
FT   CARBOHYD     29     29   POTENTIAL.
SQ   SEQUENCE    537 AA;  62024 MW;  DB5A8743 CRC32;
```

```
Query Match             51.8%;  Score 1007.5;  DB 1;  Length 537;
Best Local Similarity   55.9%;  Pred. No. 6.8e-64;
Matches  190;  Conservative  50;  Mismatches  83;  Indels  17;  Gaps   4;

Qy   20 SSEVELDCWFDEDEFKFILLPVSIAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLAL 79
         :: |||||||||||||||||||||||||| |||||||::||:|||||| ||||||||:|
Db   28 TNDTEDICVFDEGFKFILLPVSYSYSAVEMVGLPLNIAAMWIFIAKMRPWNPTIVYMFNLAL 87

Qy   80 SDTLYVLLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICH 139
         ||:|||:|||||:|||||  |||||||:||:||||||||:||:|||||||||||||||:||
Db   88 SDILYVLSLPTLVYYYADKNNMPFGEVLCKLVRFLFYANLSSILFLTCISVHRYIRGVCH 147

Qy  140 PLRALRWGRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPEEFDHVVH 199
         |:  ||| :: |   :||||:||:||::|||||:||||::||||||||||||||:||:|:
Db  148 PITSLRRMNAKHAYVICALVWLSVTLCLVPNLIFVTVSPKVYNTICHDTTRPEDFARYVE 207

Qy  200 FSSAVMGLLFGVPCLVTLVCYGLMARRLYQPL-PGSAQS--SSRLRSLRTIAVLTVFAV 256
         :||:|  ||||:|||   ||||||||:| :|| |||:|:  :|:||||:||||:|:|:|
Db  208 YSTAIMCLLFGTPCLLIAGCIGLMTRELMKPIVSGNQOTLPSYKRRSIKITLIFVMIAFAI 267

Qy  257 CFVPFHITRTIYYLARLLEADCRVLNIVNVYKVTRPLASANSC 316
         |:||||||||:||||||||  | ||||||:|||||||||||||
Db  268 CEMPFHITRTLYYARLLGIKCYALNIVINVIYKVTRPLASANSCIDPILIFFLANDRYRRR 327

Qy  317 L-------------ROLCGGGKPQ--PRTAASSLALVSLPE 342
         |              |:||||||||  |||||||||||||
Db  328 LIRTVRRSSVPNRRCMHTNHPQTEPHMTAGPLPVISAEE 367
```

RESULT 3

```
ID   P2UR_HUMAN     STANDARD;      PRT;   377 AA.
AC   P41231;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   P2U PURINOCEPTOR 1 (ATP RECEPTOR) (P2U1) (PURINERGIC RECEPTOR).
GN   P2RY2 OR P2RU1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AIRWAY EPITHELIUM;
RX   MEDLINE; 94211846.
RA   PARR C.E., SULLIVAN D.M., PARADISO A.M., LAZAROWSKI E.R., BURCH L.H.,
RA   OLSEN J.C., ERB L., WEISMAN G.A., BOUCHER R.C., TURNER J.T.;
RT   "Cloning and expression of a human P2U nucleotide receptor, a target
RT   for cystic fibrosis pharmacotherapy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3275-3279(1994).
RN   [2]
RP   REVISIONS.
RX   MEDLINE; 95108098.
RA   PARR C.E., SULLIVAN D.M., PARADISO A.M., LAZAROWSKI E.R., BURCH L.H.,
RA   OLSEN J.C., ERB L., WEISMAN G.A., BOUCHER R.C., TURNER J.T.;
RT   "Cloning and expression of a human P2U nucleotide receptor, a target
RT   for cystic fibrosis pharmacotherapy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:13067-13067(1994).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR UTP = ATP > ATP-GAMMA-S >> 2-
CC       METHYLTHIO-ATP = ADP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY
CC       G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC       MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SPLEEN, TESTIS, KIDNEY, LIVER, LUNG, HEART AND
CC       BRAIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```
```
DR   EMBL; U07225; AAC04923.1; -.
DR   EMBL; S74902; AAB32966.1; -.
DR   PIR; A54946; A54946.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1008; -.
DR   MIM; 600041; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation.
FT   DOMAIN        1     32       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     33     59       1 (POTENTIAL).
FT   DOMAIN       60     70       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     71     93       2 (POTENTIAL).
FT   DOMAIN       94    110       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    111    129       3 (POTENTIAL).
FT   DOMAIN      130    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    172       4 (POTENTIAL).
FT   DOMAIN      173    194       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    195    220       5 (POTENTIAL).
FT   DOMAIN      221    246       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    247    269       6 (POTENTIAL).
FT   DOMAIN      270    287       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    288    309       7 (POTENTIAL).
FT   DOMAIN      310    377       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      9      9       POTENTIAL.
FT   CARBOHYD     13     13       POTENTIAL.
FT   DISULFID    106    183       BY SIMILARITY.
SQ   SEQUENCE   377 AA;  42277 MW;  4C9AC77C CRC32;
```

```
Query Match
Best Local Similarity   49.6%;   Score 965;  DB 1;  Length 377;
Matches  185;  Conservative   40;  Mismatches   86;  Indels   2;  Gaps   1;

Qy   22  EVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATAYMFHLALSD  81
             ||    ||||||| ||:||:|||||||||||:||||||:|||:|:::||: ||||||:||
Db   20  ELGIRCRFNEDFKIVLLPVSYGVVCVLGLCLNAVALYIFLCRLKTWNASTTYMFHLAVSD  79

Qy   82  TLIYLSLPTLIIYYAAHNHWPFGICKFYRFLFYWNLICSVLFLTCISVHRYLGICHPL  141
             :||  |||||:|||:   |||||  || :||||||| ||| |||||||||| ||| |
Db   80  ALYAASLPLLVYYARGDHWPFSTVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPL  139

Qy  142  RALRNGRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPEEFDHYVHFS  201
             |:||  |:| :| |:|| |||:|| |:||||||||:||  | ||||:|  ||| ::|||
Db  140  RSLRWGRARYARRVAGAVWVLVLACQAPVLIFVTTSARGGRVTCHDTSAPELFSRFVAYS  199

Qy  202  SAVMGLLFGVPCLIVLVCYGLMARRLYQPLPGSAQSSSRL--RSLRTIAVVLTVFAVCFV  259
             | :|||||.||  ::|||| |||||| :|  ||:     ||  ||||||| ||:|| ||:
Db  200  SVMLGLLFAVPFAVILVCYVLMARRLLKPAYGTSGGLPRAKRKSVRTIAVVLAVFALCFL  259

Qy  260  PFHITRTIYYLARLLEADCRVLNIVNVVYKVTRPLASANSCLDPVLYILTGDKYRRQLRQ  319
             |||:||| :|  |:|  |  :|||:| :||||||||||||||||||| | || :| :| :
Db  260  PFHVTRTLIYSFRSLDLSCHTLNAINMAYKVTRPLASANSCLDPVLYFLAGQRLVRFARD  319

Qy  320  LCGGGRPQPRTAA  332
             | |    |    |
Db  320  AKPPTGPSPATPA  332
```

RESULT 4

```
ID   P2UR_MOUSE     STANDARD;      PRT;   373 AA.
AC   P35383; O09031;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   P2U PURINOCEPTOR 1 (ATP RECEPTOR) (P2U1) (PURINERGIC RECEPTOR).
GN   P2RY2 OR P2RU1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96316177.
RA   ENOMOTO K., FURUYA K., MOORE R.C., YAMAGISHI S., OKA T., MAENO T.;
RT   "Expression cloning and signal transduction pathway of P2U receptor
RT   in mammary tumor cells.";
RL   Biol. Signals 5:9-21(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93281707.
RA   LUSTIG K.D., SHIAU A.K., BRAKE A.J., JULIUS D.;
RT   "Expression cloning of an ATP receptor from mouse neuroblastoma
RT   cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5113-5117(1993).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR UTP = ATP > ATP-GAMMA-S >> 2-
CC       METHYLTHIO-ATP = ADP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY
CC       G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC       MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SPLEEN, TESTIS, KIDNEY, LIVER, LUNG, HEART AND
CC       BRAIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L14751; AAA39871.1; -.
DR   EMBL; S83099; AAB50735.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_0451; -.
DR   MGD; MGI:105107; P2RY2.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation.
FT   DOMAIN        1     32       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     33     59       1 (POTENTIAL).
FT   DOMAIN       60     70       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     71     93       2 (POTENTIAL).
FT   DOMAIN       94    110       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    111    129       3 (POTENTIAL).
FT   DOMAIN      130    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    172       4 (POTENTIAL).
FT   DOMAIN      173    194       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    195    220       5 (POTENTIAL).
FT   DOMAIN      221    246       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    247    269       6 (POTENTIAL).
FT   DOMAIN      270    287       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    288    309       7 (POTENTIAL).
FT   DOMAIN      310    373       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      9      9       POTENTIAL.
FT   CARBOHYD     13     13       POTENTIAL.
FT   DISULFID    106    183       BY SIMILARITY.
FT   CONFLICT     17     17       E -> D (IN REF. 2).
FT   CONFLICT    120    120       S -> R (IN REF. 2).
FT   CONFLICT    125    125       V -> M (IN REF. 2).
FT   CONFLICT    196    196       V -> M (IN REF. 2).
FT   CONFLICT    355    355       D -> N (IN REF. 2).
FT   CONFLICT    369    371       KDI -> PYV (IN REF. 2).
SQ   SEQUENCE   373 AA;  42174 MW;  EC153F79 CRC32;
```

```
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   P2U PURINOCEPTOR 1 (ATP RECEPTOR) (P2U1) (PURINERGIC RECEPTOR).
GN   P2RY2 OR P2RU1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95110548.
RA   RICE W.R., BURTON F.M., FIEDELDEY D.T.;
RT   "Cloning and expression of the alveolar type II cell p2u-purinergic
RT   receptor.";
RL   Am. J. Respir. Cell Mol. Biol. 12:27-32(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=PITUITARY;
RA   CHEN Z.P., KRULL N., XU S., LEVY A., LIGHTMAN S.L.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR;
RA   SEYE C.I., GADEAU A.P., DESGRANGES C.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: RECEPTOR FOR EXTRACELLULAR UTP - ATP > ATP-GAMMA-S >> 2-
    METHYLTHIO-ATP = ADP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY
    G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
    MESSENGER SYSTEM.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
    -----------------------------------------------------------------------
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
    -----------------------------------------------------------------------
DR   EMBL; U09402; AAA61565.1; -.
DR   EMBL; L46865; AAB02099.1; -.
DR   EMBL; U58839; AAC00048.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1412; -.
DR   GCRDB; GCR_1454; -.
DR   GCRDB; GCR_1512; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation.
FT   DOMAIN        1     32       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     33     59       1 (POTENTIAL).
FT   DOMAIN       60     70       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     71     93       2 (POTENTIAL).
FT   DOMAIN       94    110       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    111    129       3 (POTENTIAL).
FT   DOMAIN      130    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    172       4 (POTENTIAL).
FT   DOMAIN      173    194       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    195    220       5 (POTENTIAL).
FT   DOMAIN      221    245       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    246    268       6 (POTENTIAL).
FT   DOMAIN      269    286       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    287    308       7 (POTENTIAL).
FT   DOMAIN      309    374       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      9      9       POTENTIAL.
FT   CARBOHYD     13     13       POTENTIAL.
FT   DISULFID    106    183       BY SIMILARITY.
FT   CONFLICT    132    132       C -> S (IN REF. 1).
FT   CONFLICT    143    143       S -> R (IN REF. 1).
FT   CONFLICT    165    165       A -> V (IN REF. 1).
FT   CONFLICT    188    188       A -> D (IN REF. 1).
```

RESULT 5
ID P2UR_RAT   STANDARD;   PRT;   374 AA.
AC P41232;

Query Match 49.5%; Score 962.5; DB 1; Length 373;
Best Local Similarity 53.7%; Pred. No. 7e-61;
Matches 188; Conservative 46; Mismatches 97; Indels 19; Gaps 3;

```
Qy   22 EVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSD 81
            :||||| |:|| ||||||:| | |:| ||||||||||:|||||:||||||||
Db   20 ELGYKCRFNEDFKXVLLPVSYGVGVVCVLGLCLNVVALYIFLCRLKTWNASTTYMFHLAVSD 79

Qy   82 TLIVLSLPFLIYYYAAHHNWPFGTEICKFKVRFLFYWNLYCSVFLTCISVHRYLGICHPL 141
         :|| |||  |:|::||| :||||  :|| ||||| |||:||:|:|| || :|||:|
Db   80 SLYAASLPLLVYYYARGDHWPFSTVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPL 139

Qy  142 RALRWGRPRLAGLLCIAVWLVVAGCLVPNLFFVTTSNKGTTVLCHDTRPEEFDHYVHFS 201
          :||||| |  |:|| | |||:|||:| ||||||||:| ||||||:|||:|||:|| :
Db  140 HSLRWGRAYARRVAAVVWVLACQAPVLYFVTTSVRGTRITCHDTSARELFSHFVAYS 199

Qy  202 SAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRL--RSLRTIAVVLTVFAVCFV 259
         |: | |:| :||:|:||| ||| :||: |||: ||:||||  |:|:::::|:|:||| |:
Db  200 SVMIGLLFAVPFSVILVCVLMARRLKPAYGTGGLPRAKRKSVRTIALVLAVFALCFL 259

Qy  260 PFHITRTIYLARLLEADCRVLNIVNVVKVTRPLASNSCLDPVLYLLTGDKYRRQLRQ 319
         ||| ||:||:| ||| :| ||:|::| |||||||:|||||||||||||::||||||:
Db  260 PFHVTRTLYSFRSDLSCHTLNAINMAYKITRPLASANSCLDPVLYFLAGQRLVRFARD 319

Qy  320 LCGGGKPQ--------------RTAASSLAIVSLPEDSSCRWAATPQDS 355
          |  ||:|              |:  :|::|:: ::|||:|::|||
Db  320 AKPPTEPTPSPQARRKLGLHRPNRTVRKDL---SVSSDDSRRTESTPAGS 366
```

```
SQ   SEQUENCE   374 AA;  42137 MW;  32984BC9 CRC32;

Query Match
Best Local Similarity   48.8%;   Score 950;  DB 1;  Length 374;
Conservative  52.9%;  Pred. No. 5.2e-60;
Matches 184;         50;  Mismatches 100;  Indels  14;  Gaps   3;

Qy    22  EVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTIWLFIFRLRPWDATATYMFHLALSD  81
          |:||| |||| ||||||||:|||||||:||||||||||||:|| |||:|||||:||||||
Db    20  ELGYKCRFNEDFKIVLLPVSYGVVCVLGLCLNVVALYIFLCRLKTWNASTIYMFHLAVSD  79

Qy    82  TLIYLSLPLIIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPL  141
          :||| |||||||||| :||||||||:||:||||||:||||||||||||||||:|:|| ||
Db    80  SLYAASLPLIVIYAQGDHHPFSTVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPL  139

Qy   142  RALRWGRPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPEEFDHYVHFS  201
          ||||||||||:|||| |:|||||||||||||||:||||||||||||||||||||::|||:
Db   140  HSLSWGHARYARRVAAVVWLVLACQQAPVLYFVTSVRGTRITCHDTSARELFSHFVAYS  199

Qy   202  SAVMGLLFGVPCLVTLVCYGLMARRLYQPLPG-SAQSSSRLRSRTIAVVLTVFAVCFVP  260
          :||:||||||||||||||||:||||||||||| : ||:|||:|:|||||||:||:|||
Db   200  SVMLGLLFAVPFSIILVCYVLMARRLLKPAYGTTGLPRAKRKSVRTIALVLAVFALCFLP  259

Qy   261  FHITRTIYILYARLLEADCRVLNINVNVVYKTVTRPLASANSCLDPVLYLLTGDKYRRQLRQL  320
          ||||||||:||| |:|| | |:|:|| | :|||:||| :|||:|||||||:| |:|:||||
Db   260  FHVRTLIYSFRSLDLSCHTLNAINMAYKITRPLASANSCLDPVLYFLAGQRLVRFARDA  319

Qy   321  CGGGKPQPR-TAASSIAL--------------------VSLPEDSSCRWAATPQDS  355
          :||||||||| ||||||          ||:||||||||
Db   320  KPAEPTPSPQARRKIGLHRPNRDTVRKDLSISSDDSRRTESTPAGS  367
```

```
RESULT   6

ID   P2YR_HUMAN        STANDARD;     PRT;    373 AA.
AC   P47900;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN   P2RY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 96257237.
RA   LEON C., VIAL C., CAZENAVE J.-P., GACHET C.;
RT   "Cloning and sequencing of a human cDNA encoding endothelial P2Y1
RT   purinoceptor.";
RL   Gene 171:295-297(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96158962.
RA   AYYANATHAN K., TANIA W., HARANSJIT S., RAGHBIR A.S., BARNARD E.A.,
RA   KUNAPULI S.P.;
RT   "Cloning and chromosomal localization of the human P2Y1
RT   purinoceptor.";
RL   Biochem. Biophys. Res. Commun. 218:783-788(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96205320.
RA   JANSSENS R., COMMUNI D., PIROTTON S., SAMSON M., PARMENTIER M.,
RA   BOEYNAEMS J.M.;
RT   "Cloning and tissue distribution of the human P2Y1 receptor.";
RL   Biochem. Biophys. Res. Commun. 221:588-593(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   LEON C., VIAL C., WEBER J., CAZENAVE J.-P., GACHET C.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
```

```
CC       ATP AND ADP.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; Z49205; CAA89066.1; -.
DR   EMBL; U42030; AAA97873.1; -.
DR   EMBL; U42029; AAA97872.1; -.
DR   EMBL; S81950; AAB47091.1; -.
DR   EMBL; AJ009945; CAA07339.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1056; -.
DR   GCRDB; GCR_1884; -.
DR   GCRDB; GCR_1948; -.
DR   GCRDB; GCR_1949; -.
DR   MIM; 601167; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     52       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     53     74       1 (POTENTIAL).
FT   DOMAIN       75     87       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     88    109       2 (POTENTIAL).
FT   DOMAIN      110    126       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    127    147       3 (POTENTIAL).
FT   DOMAIN      148    166       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    167    188       4 (POTENTIAL).
FT   DOMAIN      189    218       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    219    238       5 (POTENTIAL).
FT   DOMAIN      239    265       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    266    285       6 (POTENTIAL).
FT   DOMAIN      286    303       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    304    328       7 (POTENTIAL).
FT   DOMAIN      329    373       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    124    202       BY SIMILARITY.
FT   CARBOHYD     11     11       POTENTIAL.
FT   CARBOHYD     27     27       POTENTIAL.
FT   CARBOHYD    113    113       POTENTIAL.
FT   CARBOHYD    197    197       POTENTIAL.
FT   CONFLICT    138    138       MISSING (IN REF. 1).
SQ   SEQUENCE   373 AA;  42071 MW;  4E20C20F CRC32;
```

```
Query Match
Best Local Similarity   33.0%;   Score 641.5;  DB 1;  Length 373;
Conservative  40.0%;  Pred. No. 2.1e-38;
Matches 140;         60;  Mismatches 121;  Indels  29;  Gaps   9;

Qy    17  GPGSS-----------------EVELDCWFDE-DFKFILLPVSYAVVFVLGLGLNAPTIWLFIFR  63
          |||||                 ||||||| || |||  |:||||||:||||||||||| ||
Db    20  GPGSSWGNSTVASTAAVSSSFKCALTKTGFQYYLPAVYILVFIIGFGLGNSVAIWMFVFH  79

Qy    64  LRPWDATATYMFHLALSDTLYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSV  123
          |:|||||||||||||||| |: || :|||||||||||||||:||:|||||||||||||:|
Db    80  MKPWSGISVYMFNLALADFLYVLTLPALIFYYFNKTDWIFGDAMCKLQRFIFHVNLYGSI  139

Qy   124  LFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLA--VWLVVAGCLVPNLFFVTTS-NKG  180
          |||||:||||||| |:||| |: |||:|:|| |:|   ||:||::||||:||:| || |
Db   140  LFLTCISAHRYSGVVPLKSL--GRLKKKNAICISVLWLIVVAISPILFYSGTGVRKN  197

Qy   181  TTVLCHDTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRL-YQPLPGSAQSSS  239
          |||| |||||||||||||:|::|||||||||||||||:||:|||:||| |:||||||||||
Db   198  KTITCIDTTSDEYLRSYFIYSMCTTVAMFCVPLVLIIGCYGLIVRALIYKDLD---NSPL  254

Qy   240  RLRSLRTIAVVLTVFAVCFVPFHITRTIYILYARL-----LEADCRVLNINVNVVYKVTRPLAS  296
          ||: :|| || ||||||||| ||||| |:|:||:|       |||| |||:||||:||||||
Db   255  RRKSILVIIULTVFAVSYIPFHVMKTMNLRARLDFQTPAMCAFNDRVYATYQVTRGLAS  314
```

```
OY  297  ANSCLDPVLILTGDKYRROL----ROLCGGGKPQPRTAASSLALVSLPE  342
         ||||||||||||||||:|| :| |||   |:  ||| :|||||||||||||
Db  315  LNSCVDPILYFLAGDTFRRRLSRATRKASRSEANLQSKSEDMTLNILPE  364
```

RESULT 7

```
ID   P2YR_MELGA     STANDARD;      PRT;   362 AA.
AC   P49652;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR) (6H1
DE   ORPHAN RECEPTOR).
GN   P2RY1.
OS   Meleagris gallopavo (Common turkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Meleagrididae; Meleagris.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 94335907.
RA   FILTZ T.M., LI Q., BOYER J.L., NICHOLAS R.A., HARDEN T.K.;
RT   "Expression of a cloned P2y purinergic receptor that couples to
RT   phospholipase C.";
RL   Mol. Pharmacol. 46:8-14(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97382456.
RA   LI Q., SCHACHTER J.B., HARDEN T.K., NICHOLAS R.A.;
RT   "The 6H1 orphan receptor, claimed to be the p2y5 receptor, does not
RT   mediate nucleotide-promoted second messenger responses.";
RL   Biochem. Biophys. Res. Commun. 236:455-460(1997).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC       ATP AND ADP. SEEMS TO MEDIATE ITS ACTION VIA A PERTUSSIS TOXIN
CC       INSENSITIVE G-PROTEIN, PROBABLY BELONGING TO THE GQ FAMILY THAT
CC       ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAINLY FOUND IN BLOOD, BRAIN, AND LUNG. TO A
CC       LESSER EXTENT IN STOMACH, GUT, AND SKELETAL MUSCLE.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   -------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------
DR   EMBL; U09842; AAA18784.1; -.
DR   EMBL; AF012103; AAB65428.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDb; GCR_2392; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     41  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     42     63  1 (POTENTIAL).
FT   DOMAIN       64     76  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     77     98  2 (POTENTIAL).
FT   DOMAIN       99    115  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    116    136  3 (POTENTIAL).
FT   DOMAIN      137    155  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    156    177  4 (POTENTIAL).
FT   DOMAIN      178    207  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    208    227  5 (POTENTIAL).
FT   DOMAIN      228    254  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    255    274  6 (POTENTIAL).
FT   DOMAIN      275    292  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    293    317  7 (POTENTIAL).
FT   DOMAIN      318    362  CYTOPLASMIC (POTENTIAL).
FT   DISULFID    113    191  BY SIMILARITY.
FT   CARBOHYD     11     11  POTENTIAL.
FT   CARBOHYD     26     26  POTENTIAL.
FT   CARBOHYD    102    102  POTENTIAL.
FT   CARBOHYD    186    186  POTENTIAL.
SQ   SEQUENCE   362 AA;  41180 MW;  B9D63032 CRC32;
```

```
Query Match   32.4%;  Score 631;  DB 1;  Length 362;
Best Local Similarity   39.7%;  Pred. No. 1.1e-37;
Matches  138;  Conservative  61;  Mismatches 133;  Indels 16;  Gaps 7;

Qy    4  TESSLLRSLGLSLSPGPGSSEVELDCWFDEDEKFILLPVSYAVVFVLGLGNAPTLWLFIFR  63
         |: ||   |  |    :|   : |::       :  |||:||:|||  |||:| : :|:|:|
Db   13  TOPELLAG-GWAAGNASTKCSLT---KTGFQYYLPTVYILVFIITGFLGNSVAIWMFVFH  68

Qy   64  LRPWDATATYMFHLALSDTILYVLSLPTLIIYYAAHNHWPFGTEICKFVRFLFYWNLYCSV  123
         ||| |:|:|:| |:|||||||:||:|||| || |:||:||| |:|:|:|:| |||| |||:|
Db   69  MRPWSGISVYMFNLALADFLYVLTLPALIFYYFNKTDWIFGDVMCKLQRFIFHVNLYGSI  128

Qy  124  LFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVWLVVAGCLVPNLFVTTS-NKGTT  182
         :|||||||:|||||| |:| : |   | | || |: : |  | :|  ||| : ||| |
Db  129  LFLTCISVHRYTGVVHPLKSLGRLKKKNAVVSSLVWALVVAIRAPILFYSGTGVRRNKT  188

Qy  183  VLCHDTTRPEEFDHVVHFESSAVMGLLFGYPCLVLVCYGLMARRL-YOPLPGSAOSSSRL  241
         |: |:  | :: ||:  | |: |   ||| |: |: |: |:|  | :  || :   |:
Db  189  ITCVDTTADEYLRSFVVSMCTTVPMFCIPFIVILGCYGLIVKALIYKDLD---NSPLRR  245

Qy  242  RSLRTIAVVLTVFAVCFVPFHITRRIIYILARL---LEADCRVLNIVNVVKVTRPLASAN  298
         |:| :||||||:| :  || || |  | | ||   ||  |  |: | |:| :||  | |
Db  246  KSILVIVILTVFAVSYLPFHVMKTLNLRARLDFQTPOMCAFNDKVYATYQVTRGLASLN  305

Qy  299  SCLDPVLILTGDKYRROL----ROLCGGGKPQPRTAASSLALVSLPE  342
         |||||| | || |||||    | | |||   |  | ::||||:|| :
Db  306  SCVDPILYFLAGDTFRRRLSRATRKASRSEPNVQSKSEEMTLNILTE  353
```

RESULT 8

```
ID   P2YR_CHICK     STANDARD;      PRT;   362 AA.
AC   P34996;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN   P2RY1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 93285340.
RA   WEBB T.E., SIMON J., KRISHEK B.J., BATESON A.N., SMART T.G.,
RA   KING B.F., BURNSTOCK G., BARNARD E.A.;
RT   "Cloning and functional expression of a brain G-protein-coupled ATP
RT   receptor.";
RL   FEBS Lett. 324:219-225(1993).
RN   [2]
RP   3D-STRUCTURE MODELING.
RX   MEDLINE; 97026278.
RA   VAN RHEE A.M., FISCHER B., VAN GALEN P.J.M., JACOBSON K.A.;
RT   "Modelling the P2Y purinoceptor using rhodopsin as template.";
RL   Drug Des. Discov. 13:133-140(1995).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC       ATP AND ADP. SEEMS TO MEDIATE ITS ACTION VIA A PERTUSSIS TOXIN
CC       INSENSITIVE G-PROTEIN, PROBABLY BELONGING TO THE GQ FAMILY THAT
CC       ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: BRAIN, SPINAL CORD, GASTROINTESTINAL TRACT,
CC       SPLEEN AND LEG MUSCLE. IS NOT DETECTED IN THE HEART, LIVER,
CC       STOMACH, LUNG AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X73268; CAA51716.1; -.
DR   PIR; S33733; S33733.
DR   PDB; 1DDD; 11-JUL-96.
DR   GCRDB; GCR_0745; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; 3D-structure.
FT   DOMAIN        1     41       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     42     63       1 (POTENTIAL).
FT   DOMAIN       64     76       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     77     98       2 (POTENTIAL).
FT   DOMAIN       99    115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    116    136       3 (POTENTIAL).
FT   DOMAIN      137    155       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    156    177       4 (POTENTIAL).
FT   DOMAIN      178    207       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    208    227       5 (POTENTIAL).
FT   DOMAIN      228    254       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    255    274       6 (POTENTIAL).
FT   DOMAIN      275    292       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    293    317       7 (POTENTIAL).
FT   DOMAIN      318    362       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    113    191       BY SIMILARITY.
FT   CARBOHYD     11     11       POTENTIAL.
FT   CARBOHYD     26     26       POTENTIAL.
FT   CARBOHYD    102    102       POTENTIAL.
FT   CARBOHYD    186    186       POTENTIAL.
SQ   SEQUENCE   362 AA;  41194 MW;  3713AE31 CRC32;

Query Match             32.3%;
Best Local Similarity   39.4%;          Score 628;  DB 1;  Length 362;
Matches   137;  Conservative   62;  Mismatches 133;  Indels  16;  Gaps   7;   Pred. No. 1.8e-37;

Qy     4  TESSLLRSLGLSPGPGSSEVELDCWFDEDEFKILLPVSYAVVFVLGLGLNAPTLWLFIFR   63
          :| :|||||  ::|         |      |:|:| ||||||||  |||||||||||||||
Db    13  TQPELLAG-GWAAGNATTKCSLT---KTGFQFYYLPTVYILVFITGFLGNSVAIWMFVFH   68

Qy    64  LRPWDATATYMFHLALSDTLYVLSLPTLIYYYAAHNHWPFGTEICKFYRFLYWNLYCSV  123
          :|||  ::::||:||||| ||||:||:|::|:  ::  |:|  ||| ||::||||| |:
Db    69  MRPWSGISVYMFNLALADFLYVLTLPALIFYFENKTDWIFGDVMCKLQRFIFHVNLYGSI  128

Qy   124  LFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVWLVVAGCLVPNLFFVVTS-NKGTT  182
          |||||||||||:|::|||::| ::   ::: :| |||||: || | ||:| | | |:::|
Db   129  LFLTCISVHRYTGVVHPLKSLGRLKKKNAVYVSSLVWALVVAVIAPILFYSGTGVRRNKT  188

Qy   183  VLCHDTTRPEEFDHYVHRSSAVNGLLGVPCLVTLVCYGLMARRL-YQPLPGSAQSSSRL  241
          : |:|||  ::: :|: || |  |:|  || |::||||||:: || :  |  :  ::| |
Db   189  ITCYDTTADEYLRSYFVYSMCTTVFMFCIPFIVLGCYGLIVKALIYKDLD---NSPLRR  245

Qy   242  RSLRTIAVVLTVFAVCFVPFHITRTIYILARL---LEADCRVLNIVNVVKVTRPLASAN  298
          :|: ::  ||:| || ::|||: ||:::  ||       | :| :::  |:|:|:|::|
Db   246  KSIYLVIVLTVPAVSVLPFHVNKTLNLRARLDFQTPQMCAFNDKVIATYQVTRGLASLN  305

Qy   299  SCLDPVLYLLTGDKYRRQL----RQLCGGGKPQPRTASSLALVSLPE  342
          || ||:||:| || ::|:|        ::| | ::::|:::|:: :|
Db   306  SCVDPILYFLAGDTFRRRLSRATRKSSRRSEPNVQSKSEEMTLNILTE  353
```

```
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN   P2RY1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AORTIC ENDOTHELIUM;
RX   MEDLINE; 95352058.
RA   HENDERSON D.J., ELLIOT D.G., SMITH G.M., WEBB T.E., DAINTY I.A.;
RT   "Cloning and characterisation of a bovine P2Y receptor.";
RL   Biochem. Biophys. Res. Commun. 212:648-656(1995).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC       ATP AND ADP.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X87628; CAA60958.1; -.
DR   HSSP; P34996; 1DDD.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
FT   DOMAIN        1     52       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     53     74       1 (POTENTIAL).
FT   DOMAIN       75     87       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     88    109       2 (POTENTIAL).
FT   DOMAIN      110    126       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    127    147       3 (POTENTIAL).
FT   DOMAIN      148    166       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    167    188       4 (POTENTIAL).
FT   DOMAIN      189    218       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    219    238       5 (POTENTIAL).
FT   DOMAIN      239    265       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    266    285       6 (POTENTIAL).
FT   DOMAIN      286    303       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    304    328       7 (POTENTIAL).
FT   DOMAIN      329    373       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    124    202       BY SIMILARITY.
FT   CARBOHYD     11     11       POTENTIAL.
FT   CARBOHYD     27     27       POTENTIAL.
FT   CARBOHYD    113    113       POTENTIAL.
FT   CARBOHYD    197    197       POTENTIAL.
SQ   SEQUENCE   373 AA;  42287 MW;  5C9B7B5 CRC32;

Query Match             31.9%;
Best Local Similarity   43.4%;          Score 621;  DB 1;  Length 373;
Matches   126;  Conservative   53;  Mismatches 103;  Indels   8;  Gaps   4;

Qy    33  FKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLI   92
          |:| ||||||||||||||||||||||||||:|:||||||||||||||||||||||||||
Db    49  FQFYYLPAVYILVIFIIGFLGNSVAIWMFVFHMKPWSGISVYMFNLALADFLYVLTLPALI  108

Qy    93  YYYAAHNHWPFGTEICKFYRFLYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLA  152
          :|:  ::  |:|  ||| ||::||||| |: |||||||||||:|::||||:|| ::: :|
Db   109  FYFNKTDWIFGDAMCKLQRFIFHVNLYGSILFLTCISAHRYSGVVHPLKSLGRLKKKNA  168

Qy   153  GLLCLAVWLVVAGCLVPNLFFVVTS-NKGTTVLCHDTTRPEEFDHYVHRSSAVNGLLFGV  211
          |::  :|||:|| || ||:| | | |:::| : |:||  ::: :|: || |  |:|   |
Db   169  VYISVLVWLLVVGISPILFYSGTGIRKNKTITCIDTTSDEYLRSYFIYSMCTTVAMFCV  228

Qy   212  PCLVTLVCYGLMARRL-YQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYYL  270
Db   169  VYISVLVWLLVVGISPILFYSGTGIRKNKTITCIDTTSDEYLRSYFIYSMCTTVAMFCV  270
```

RESULT 10

```
Db
Qy   271 ARL---LEADCRVLNIVNVVKVTRPLASANSCLDPVLILLTGDKYRRQL 317
Db   229 PLVLILGCYGLIVRALIYKDLD---NSPLRRKSIYLIVLIVLTVPAVSYIPFHVWKTMNLR 285
     286 ARLDFQTPEMCAFNDRYATYQVTRGLASLNSCVDPILYFLAGDTFRRL 335
```

```
ID   P2YR_RAT     STANDARD;     PRT;   373 AA.
AC   P49651;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN   P2RY1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INSULINOMA;
RX   MEDLINE; 95298025.
RA   TOKUYAMA Y., HARA M., JONES E.M.C., FAN Z., BELL G.I.;
RL   "Cloning of rat and mouse P2Y purinoceptors.";
RL   Biochem. Biophys. Res. Commun. 211:211-218(1995).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC       ATP AND ADP. IN PANCREATIC ISLETS, MAY MEDIATE SOME OF THE EFFECTS
CC       OF EXTRACELLULAR ATP ON INSULIN SECRETION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MUSCLE, HEART, LIVER, KIDNEY,
CC       LUNG, BRAIN, SPLEEN, BUT NOT IN TESTIS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
DR   EMBL; U22830; AAA91303.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1444; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN         1     52   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      53     74   1 (POTENTIAL).
FT   DOMAIN        75     87   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      88    109   2 (POTENTIAL).
FT   DOMAIN       110    126   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     127    147   3 (POTENTIAL).
FT   DOMAIN       148    166   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     167    188   4 (POTENTIAL).
FT   DOMAIN       189    218   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     219    238   5 (POTENTIAL).
FT   DOMAIN       239    265   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     266    285   6 (POTENTIAL).
FT   DOMAIN       286    303   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     304    328   7 (POTENTIAL).
FT   DOMAIN       329    373   CYTOPLASMIC (POTENTIAL).
FT   DISULFID     124    202   BY SIMILARITY.
FT   CARBOHYD      11     11   POTENTIAL.
FT   CARBOHYD      27     27   POTENTIAL.
FT   CARBOHYD     113    113   POTENTIAL.
FT   CARBOHYD     197    197   POTENTIAL.
SQ   SEQUENCE   373 AA;   42321 MW;  FAF215F8 CRC32;
```

Query Match   31.7%;   Score 616;   DB 1;   Length 373;

RESULT 11

```
Best Local Similarity 43.1%;   Pred. No. 1.3e-36;
Matches 125;  Conservative 54;  Mismatches 103;  Indels 8;  Gaps 4;

Qy    33 FKFILLPVSIAVFVLGLLGLNAPFLWLFIFRLRPWDARATYMFHLALSDTLYVLSLPTLI   92
Db       |:|:||||| |:|||||||||:|||||:|||||||:|||:||||:|||||:||||:
      49 FQFYLPAVYILVFIIGFLGNSVAIWMFVFHMKPWSGISVYMFNLALADFIYVLTLPALI  108

Qy    93 YYYAAHNHHPFGEICKFVRFLFYWNLLCGVSLFLTCISVHRYIGICHPLRALRWGRPRLA  152
Db       |||:| || ||||| ||||:||||:|:|||||||||:|||:||||||:||||
     109 FYYFNKTDWIFGDVMCKLQRFIFHVNLYGSISLFLTCISAHRYSGVVYPLKSLGRLKKKNA  168

Qy   153 GLLCLAVWLVVAGCLVPNLFFVTTS-NKGTTVLCHDTTRPEEFDHIVHFSSAVMGLLFGV  211
Db       |||:||||:|||||||||||| |||||:|||||||:|:|:|:||||||||
     169 IYVSVLVWLIVVVAISPILFYSGTGIRKNKTVTCYDSTSDEYLRSYFIYSMCTTVAMFCI  228

Qy   212 PCLVTLVCYGLMARRL-YQPLPGGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYYL  270
Db       ||||||||||:||||:||||||:|:|||||:|||||||||||||||:|:|||||
     229 PLVLILGCYGLIVRALIYKDLD---NSPLRRKSIYLIVLIVLTVPAVSYIPFHVWKTMNLR  285

Qy   271 ARL---LEADCRVLNIVNVVKVTRPLASANSCLDPVLILLTGDKYRRQL  317
Db       ||||||||||:||||||||||||||||||:|:|||||:|||:||:
     286 ARLDFQTPEMCAFNDRYATYQVTRGLASLNSCVDPILYFLAGDTFRRL  335
```

```
ID   P2YR_MOUSE     STANDARD;     PRT;   373 AA.
AC   P49650;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN   P2RY1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INSULINOMA;
RX   MEDLINE; 95298025.
RA   TOKUYAMA Y., HARA M., JONES E.M.C., FAN Z., BELL G.I.;
RL   "Cloning of rat and mouse P2Y purinoceptors.";
RL   Biochem. Biophys. Res. Commun. 211:211-218(1995).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC       ATP AND ADP. IN PANCREATIC ISLETS, MAY MEDIATE SOME OF THE EFFECTS
CC       OF EXTRACELLULAR ATP ON INSULIN SECRETION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
DR   EMBL; U22829; AAA91302.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1703; -.
DR   MGD; MGI:105049; P2RY1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN         1     52   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      53     74   1 (POTENTIAL).
FT   DOMAIN        75     87   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      88    109   2 (POTENTIAL).
FT   DOMAIN       110    126   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     127    147   3 (POTENTIAL).
FT   DOMAIN       148    166   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     167    188   4 (POTENTIAL).
FT   DOMAIN       189    218   EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM    219    238       5 (POTENTIAL).
FT   DOMAIN      239    265       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    266    285       6 (POTENTIAL).
FT   DOMAIN      286    303       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    304    328       7 (POTENTIAL).
FT   DOMAIN      329    373       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    124    202       BY SIMILARITY.
FT   CARBOHYD     11     11       POTENTIAL.
FT   CARBOHYD     27     27       POTENTIAL.
FT   CARBOHYD    113    113       POTENTIAL.
FT   CARBOHYD    197    197       POTENTIAL.
SQ   SEQUENCE    373 AA;  42212 MW;  C699E593 CRC32;

Query Match            31.6%;  Score 614;  DB 1;  Length 373;
Best Local Similarity  42.8%;  Pred. No. 1.8e-36;
Matches 124;  Conservative  55;  Mismatches 103;  Indels  8;  Gaps  4;

Qy   33  FKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLI  92
             |:|:| | |||:| ||||:||
Db   49  FQFYLPAVYILVFIIGFLGNSVAIMWFVFHMKPWSGISVTMFNLALADFLIVLTLPALI  108

Qy   93  YYYAAHNHWPFGTEICKFRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLA  152
Db  109  FYIFNKTDWIFGDAMCKLQRFIFHVNLYGSILFLTCISAHRYSGVVFPLKSLGRLKKKNA  168

Qy  153  GLLCLAVWLVVAGCLVPNLFFV-TTSNKGTTVLCHDTTRPEEFDHYVHFSSAVMGLLFGV  211
Db  169  IYVSLVWLIVVVAISPILFYSGTGVRKNKVTCYDTTSNDYLRSYFIYSMCTTVAMFCI  228

Qy  212  PCLVLTLVCYGLMARRL-YQPLPGSAQSSSRLRSLRTIAVVLLTVFAVCVPFHITRTIYYL  270
Db  229  PLVILLGCYGLIVKALIYNDLD---NSPLRRKSIYLVIVILVIVFAVSYIPFHVMKTMNLR  285

Qy  271  ARL---LEADCRVLNIVNVVKVTRPLASANSCLLDPVLYLYLTGDKYRRQL  317
Db  286  ARLDFQTPEMCDFNDRVYATYQVTRGLASLNSCVDPILYFLAGDTFRRRL  335


RESULT 12
P2Y3_CHICK

ID   P2Y3_CHICK     STANDARD;      PRT;   328 AA.
AC   Q98907;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   P2Y PURINOCEPTOR 3 (P2Y3) (NUCLEOSIDE DIPHOSPHATE RECEPTOR).
GN   P2RY3.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 96319774.
RA   WEBB T.E., HENDERSON D., KING B.F., WANG S., SIMON J.,
RA   BATESON A.N., BURNSTOCK G., BARNARD E.A.;
RT   "A novel G protein-coupled P2 purinoceptor (P2Y3) activated
RT   preferentially by nucleoside diphosphates.";
RL   Mol. Pharmacol. 50:258-265(1996).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADP > UTP > ATP = UDP. THE
CC       ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC       A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
CC   --------------------------------------------------------------------------
DR   EMBL; X92283; CAA63930.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDb; GCR_1103; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; False_neg.
FT   TRANSMEM     23     43       1 (POTENTIAL).
FT   TRANSMEM     58     78       2 (POTENTIAL).
FT   TRANSMEM    140    160       3 (POTENTIAL).
FT   TRANSMEM    185    205       4 (POTENTIAL).
FT   TRANSMEM    232    252       5 (POTENTIAL).
FT   DISULFID     94    172       BY SIMILARITY.
FT   CARBOHYD      5      5       POTENTIAL.
SQ   SEQUENCE    328 AA;  37586 MW;  8EDBC911 CRC32;

Query Match            31.0%;  Score 603;  DB 1;  Length 328;
Best Local Similarity  42.3%;  Pred. No. 9.2e-36;
Matches 127;  Conservative  45;  Mismatches 120;  Indels  8;  Gaps  4;

Qy   27  CWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVL  86
Db   13  CTFHEEFKQVLLPVYSVVFLLGLPLNAVVIGQIWLARKALTRTTIYMLNLAMADLLYVC  72

Qy   87  SLPLLIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRW  146
Db   73  SLPLLIIYNTYQKDWPFGDTCKVFRFQFYTNLHGSILFLTCISVQRHVGICHPLAS--W  130

Qy  147  GR---PRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPEEFDHYVHFSSA  203
Db  131  HKKKGKKLIWLVCAAVWFIVIAQCLPTFVFASIGTQRNRVCYDLSPPDRSTSYFPYGIT  190

Qy  204  VMGLLFGVPCLVTLVCYGLMARRLYQ--PLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPF  261
Db  191  LTIIGFLLPFAAILACYCSMARICQKDELIGLAVHKKKDKAVRMIIIVVIVFSISFFPF  250

Qy  262  HITRTIYYLAR-LLEADCRVLNIVNVVKVTRPLASANSCLLDPVLYLYLTGDKYRRQLRQL  320
Db  251  HLTKIYILIVRSSASLPCPTLQAFAIAYKCTRPFASMNSVLDPILYFFTQRKFRESTRYL  310


RESULT 13
P2Y6_RAT

ID   P2Y6_RAT       STANDARD;      PRT;   328 AA.
AC   Q63371;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   P2Y PURINOCEPTOR 6 (P2Y6).
GN   P2RY6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AORTIC SMOOTH MUSCLE;
RX   MEDLINE; 96064682.
RA   CHANG K., HANAOKA K., KUMADA M., TAKUWA Y.;
RT   "Molecular cloning and functional analysis of a novel P2 nucleotide
RT   receptor.";
RL   J. Biol. Chem. 270:26152-26158(1995).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR UTP > ADP = 2-METHYLTHIO-ATP
CC       > ADP-BETA-S > ATP = ATP-GAMMA-S. THE ACTIVITY OF THIS RECEPTOR IS
CC       MEDIATED BY G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-
CC       CALCIUM SECOND MESSENGER SYSTEM. FUNCTIONALLY COUPLED TO
CC       PHOSPHOLIPASE C.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN VARIOUS TISSUES
CC       INCLUDING LUNG, STOMACH, INTESTINE, SPLEEN, MESENTERY, HEART, AND,
CC       MOST PROMINENTLY, AORTA.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

Left column

DR   EMBL; D63665; BAA09816.1; -.
DR   HSSP; P34996; 1DDD.
DR   GCRDB; GCR_1520; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN   1   27   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   28   48   1 (POTENTIAL).
FT   DOMAIN   49   62   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   63   83   2 (POTENTIAL).
FT   DOMAIN   84   101   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   102   122   3 (POTENTIAL).
FT   DOMAIN   123   144   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   145   165   4 (POTENTIAL).
FT   DOMAIN   166   194   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   195   215   5 (POTENTIAL).
FT   DOMAIN   216   236   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   237   257   6 (POTENTIAL).
FT   DOMAIN   258   280   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   281   303   7 (POTENTIAL).
FT   DOMAIN   304   328   EXTRACELLULAR (POTENTIAL).
FT   DISULFID   99   177   BY SIMILARITY.
FT   CARBOHYD   5   5   POTENTIAL.
SQ   SEQUENCE   328 AA; 36677 MW; 2FA27E5C CRC32;

Query Match 30.2%; Score 588; DB 1; Length 328;
Best Local Similarity 42.2%; Pred. No. 1e-34;
Matches 136; Conservative 37; Mismatches 133; Indels 16; Gaps 6;

```
Qy    5 ESSLLRSLGLSPGPGSSEVELDCWFEDEDFKFILLPVSIAVFVLGLNAPTLMLFIFRL  64
                           :||:|||||||::||| |||| ::|| | |||| ||
Db    4 ----------TTCVYREDFKRILLPPVISVVLVVGLPLNVCVIAQICASR  55

Qy   65 RPWDATATYMFHLALSDTLYVLSLPTLIYYAAHNHWPFGTEICKFVRFLFYNNLYCSVL  124
         : |:::|| ||:|||||:|| |||||:|| |||||| |  ::||||||||:|| |:||
Db   56 RQLTRSAVYTLNLALADLLYACSLPLIIYNYARGDHWPFGDLACRLVRFLFYANLHGSIL  115

Qy  125 FLTCISVHRYLGICHPLRALRW---GRPRLAGLLCLAVWLVVAGCLVPNLFVTVISNKGT  181
         ||||||:|||||||||| |     :||   ||||| |||:|| || :| | || | |
Db  116 FLTCISFORYLGICHPLAP--WHKRGGRRAAWVVCGVVWLVVVTAQCLPTAVFAATGIQRN  173

Qy  182 TVLCHDTTRPEEFDHYVHFSSAVNGLLFGVPCLVTLVCYGLMARRL-YOPLP-GSAOSSS  239
         : |: ||  |:|:|  |::|  ||| |||:| |||:: | |:|||| | || |
Db  174 RTVCYDLSPPILSTRYLPYGMALTVIGFLLPFTALLACYCRMARRLCRQDGPAGPVAQER  233

Qy  240 RLRSLRTIAVVLIVFAVCFVPFHITRTIYYLARLLE-ADCRYLNIVINVNVYKVTRPLASAN  298
         :|:|| ||:| |:||||:|:|||| | |||:|:| | |||| |  :| |: |:|||:||||
Db  234 RSKAARMAVVVAAVFVISFLPFHITKTAYLAVRSTPGVSGCPVLETFAAAYKGTRPFASAN  293

Qy  299 SCLDPVLYLLTGDKYRQLRQL  320
         | |:|:|||||| |:|:|:|
Db  294 SVLDPILFYFTQOKFRQPHDL  315
```

RESULT 14

ID   P2Y6_HUMAN   STANDARD;   PRT;   328 AA.
AC   O15077; Q15754;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   P2Y PURINOCEPTOR 6 (P2Y6).
GN   P2RY6.
OS   Homo sapiens (Human).

Right column

OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   COMMUNI D., PARMENTIER M., BOEYNAEMS J.M.;
RA   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97432828.
RA   SOMERS G.R., HAMMET F., WOOLLATT E., RICHARDS R.I., SOUTHEY M.C.,
RA   VENTER D.J.;
RT   "Chromosomal localization of the human P2y6 purinoceptor gene and
RT   phylogenetic analysis of the P2y purinoceptor family.";
RL   Genomics 44:127-130(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 97415792.
RA   MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
RT   "Cloning of P2Y6 cDNAs and identification of a pseudogene: comparison
RT   of P2Y receptor subtype expression in bone and brain tissues.";
RL   Biochem. Biophys. Res. Commun. 237:297-302(1997).
RN   [4]
ERRATUM.
RX   MEDLINE; 98069816.
RA   MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
RL   Biochem. Biophys. Res. Commun. 240:298-302(1997).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR UDP > UTP > ATP. THE ACTIVITY
CC   OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE A
CC   PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --
DR   EMBL; X97058; CAA65770.1; -.
DR   EMBL; U52464; AAB03572.1; -.
DR   EMBL; AF007892; AAB80713.1; -.
DR   EMBL; AF007891; AAB80712.1; -.
DR   HSSP; P34996; 1DDD.
DR   MIM; 602451; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN   1   27   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   28   48   1 (POTENTIAL).
FT   DOMAIN   49   62   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   63   83   2 (POTENTIAL).
FT   DOMAIN   84   101   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   102   122   3 (POTENTIAL).
FT   DOMAIN   123   144   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   145   165   4 (POTENTIAL).
FT   DOMAIN   166   194   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   195   215   5 (POTENTIAL).
FT   DOMAIN   216   236   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   237   257   6 (POTENTIAL).
FT   DOMAIN   258   280   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   281   303   7 (POTENTIAL).
FT   DOMAIN   304   328   EXTRACELLULAR (POTENTIAL).
FT   DISULFID   99   177   BY SIMILARITY.
FT   CARBOHYD   5   5   POTENTIAL.
FT   CONFLICT   2   3   MISSING (IN REF. 2).
SQ   SEQUENCE   328 AA; 36429 MW; 29C125CE CRC32;

Query Match 30.1%; Score 586; DB 1; Length 328;
Best Local Similarity 42.5%; Pred. No. 1.4e-34;

Matches 135; Conservative 37; Mismatches 128; Indels 18; Gaps 6;

```
Qy     10 RSLGLSPGPGSSEVELDCWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDA 69
             ::|||  |      |  ||||||:||::|| :|||  |  ||:||||||||:| :
Db      9 QALGLP--------TTCVYRENFKOLLLPVYSAVLAAGLPLNICIVITQICTSRRALTR 60

Qy     70 TATYMFHLALSDTLYVLSLPLIIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCI 129
             |||:|||:|||||||||:||||||||| |||| ::|||| ||:|| | ||||||||||
Db     61 TAVYTLNIALADLLYACSLPLLIYNYAQGDHWPFGDPACRLVRFLFYANLHGSILFLTCI 120

Qy    130 SVHRYLGICHPLRALRW---GRPRLAGLLCLAAWLVVAGCLVPNLFVTTSNKGTTVLCH 186
             | |||||||||||||||    ||||||||||| |||||| ||||||||||||||:||||
Db    121 SFQRYLGICHPLAP--WHKRGGRRAAWLVCVAVWLAVTTQCLPTAIFAATGIQRNRVCY 178

Qy    187 DTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQ----PLPGSAQSSSRLRS 243
             |::  ||||||||||||| | | ||||||||||||||| |:| |
Db    179 DLSPPALATHYMPYGMALTVIGFLLPFAALLACYCLLACRLCRQDGPAEPVAQ-ERRGKA 237

Qy    244 LRTIAVVLTVFAVCFVPFHITRTIYIYLARLLE-ADCRVLNIVNVVKVTRPLASANSCLD 302
             ||:|  :||||||| |||||| ||| | |  ||  || || || || | || |||||| |
Db    238 ARMAVVVAAAFAISFLPFHIRKTAYLAVRSTPGVPCTVLEAFAAAYKGTRPFASANSVLD 297

Qy    303 PVLYLLITGDKYRRQLRQL 320
             ||:||| |||:||:| :|
Db    298 PILFYFTQKKFRRRPHEL 315
```

RESULT 15
GPRH_HUMAN

Query Match   26.2%; Score 510; DB 1; Length 339;
Best Local Similarity   37.1%; Pred. No. 3.1e-29;
Matches 119; Conservative 52; Mismatches 136; Indels 14; Gaps 7;

```
Qy     27 CWFDEDFKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDADATATYMFHLALSDTLYVL 86
             | : ::| |||||| :||| |||| |||:||||||||:| :||:::| ||||:|||||||||
Db     23 CGQETPLENMLFASFYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVL 82

Qy     87 SLPLIIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRW 146
             ||||:|| | |||||  | |   ||| ||: |:|||||||| || || |||||||||||
Db     83 VLPTRLVYHFSGNHWPFGEIACRLTGFLYLNMYASIYFLTCISADRFLAIVHPVKSLKL 142

Qy    147 GRPRLAGLLCLAAWLVVAGCLVPNLFVTTSNKGTTVLCHDTTRPEEFDHYVHFSSAVMG 206
             ||:|| || |  |  :|| ||| ||||  | | |  |  ||||||||||||||||||||
Db    143 RRPLYAHLACAFLWVVAVAMAPLLVSPQTVQIHHTVCLQLYR-EKASHHALVSLAV-- 200

Qy    207 LLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHITRT 266
             |||  ||||||||||||||||||| :| | :|| ||||||||||||||||||| |:|||
Db    200 -AFTPFFITTVTCYLLIIRSLRDGL--RVEKRLKTKAVRMIAIVLAIFLVCFVPYHVNRS 256

Qy    267 IYYL-ARLLEADCRVLNIVNVVKVTRPLASANSCLDPVLYLLITGDKYRRQL-RQLCG-- 323
             ||:| |::|| | |  |::|:||||||||||||| | ||:|:| |:||||| ||||
Db    257 VVYLHYRSHGASCATORILALANRITSCLTSLNGALDPIMYFVVAEKRHALGNLLCGKR 316

Qy    323 --GGKP--QPRTAASSLALVS 339
               ::| :  | ||::| |:| |
Db    317 LKGPPPSFEGKINESSLSAKS 337
```

Search completed: December 29, 1999, 20:19:11
Job time: 2130 sec

```
OM protein - protein search, using sw model

                                          GenCore version 4.5
                               Copyright (c) 1993 - 1998 Compugen Ltd.

Run on:            December 30, 1999, 20:41:04 ; Search time 16.07 Seconds
                                                 (without alignments)
                                                 1391.297 Million cell updates/sec

Title:          US-09-077-173a-2
Perfect score:  1946
Sequence:       1 MASTESSLLRSLGLSPGPGS...........CRWAATPQDSSCSTPRADRF 365
Scoring table:  BLOSUM62
Searched:       199794 seqs, 61255205 residues
Database :      SPTREMBL_11:*

Word size :     0

Number of hits that pass the threshold :   199794

                  1: sp_archea:*
                  2: sp_bacteria:*
                  3: sp_fungi1:*
                  4: sp_human:*
                  5: sp_invertebrate:*
                  6: sp_mammal:*
                  7: sp_mhc:*
                  8: sp_organelle:*
                  9: sp_phage:*
                 10: sp_plant:*
                 11: sp_rodent:*
                 12: sp_virus:*
                 13: sp_vertebrate:*
                 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                        SUMMARIES

                            %
Result             Query
No.      Score     Match  Length  DB  ID        Description
-------  --------  -----  ------  --  -------   -----------
1        1593      81.9   361     11  035811    035811 rattus norv
2        1127.5    57.9   374     11  057466    057466 meleagris g
3        599       30.8   328     13  093361    093361 meleagris g
4        450.5     23.2   370     4   015132    015132 homo sapien
5        400       20.6   353     13  P79960    P79960 xenopus lae
6        374       19.2   415     13  015185    015185 homo sapien
7        368       18.5   383     13  042324    042324 catostomus
8        356.5     18.3   359     6   077590    077590 ovis aries
9        354       18.2   339     4   000398    000398 homo sapien
10       352       18.1   303     13  P70058    P70058 xenopus lae
11       352       18.1   373     13  057585    057585 brachydanio
12       351       18.0   383     12  089609    089609 equine herp
13       349       17.9   133     6   018951    018951 bos taurus
14       348.5     17.9   381     4   095853    095853 homo sapien
15       342       17.6   371     4   043190    043190 homo sapien
16       337.5     17.3   380     6   097666    097666 macaca mula
17       334       17.2   342     13  093239    093239 cyprinus ca
18       332       17.1   356     6   097665    097665 macaca mula
19       332       17.1   353     13  093247    093247 cyprinus ca
20       331.5     17.0   396     13  088634    088634 mus musculu
21       329.5     16.9   358     13  Q9yGC3    Q9yGC3 xenopus lae
22       328       16.9   352     6   077488    077488 cercopithec
23       328       16.9   357     13  042445    042445 oncorhynchu
24       327.5     16.8   360     4   060835    060835 homo sapien
25       325       16.7   385     4   076067    076067 homo sapien
```

```
                                        ALIGNMENTS

26       324.5     16.7   356     6   097571    097571 canis famil
27       324       16.7   382     11  009027    009027 rattus norv
28       321.5     16.5   363     11  Q92026    Q92026 meriones un
29       321.5     16.5   372     13  093237    093237 cyprinus ca
30       320       16.4   362     11  089039    089039 rattus norv
31       319.5     16.4   355     6   097664    097664 macaca mula
32       319       16.4   337     4   075819    075819 homo sapien
33       317       16.3   368     13  042444    042444 oncorhynchu
34       315.5     16.2   309     4   043495    043495 homo sapien
35       315.5     16.2   359     11  054814    054814 rattus norv
36       314       16.1   378     11  008707    008707 mus musculu
37       313.5     16.1   367     11  008410    008410 mus musculu
38       313       16.1   353     11  035797    035797 rattus norv
39       312       16.0   367     6   046685    046685 bos taurus
40       308.5     15.9   367     4   054689    054689 mus musculu
41       308.5     15.9   358     13  Q92213    Q92213 cavia porce
42       306       15.7   384     4   000590    000590 homo sapien
43       306       15.7   360     11  018793    018793 macaca mula
44       306       15.7   382     11  Q92282    Q92282 mus musculu
45       305       15.7   384     4   000537    000537 homo sapien
```

```
RESULT   1

035811

ID      035811                   PRELIMINARY;     PRT;     361 AA.
AC      035811;
DT      01-JAN-1998 (TrEMBLrel. 05, Created)
DT      01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT      01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE      G-PROTEIN COUPLED RECEPTOR.
GN      P2Y4.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX      MEDLINE; 98421785.
RA      WEBB T.E., HENDERSON D., ROBERTS J.A., BARNARD E.A.;
RT      "Molecular cloning and characterization of the rat P2Y4 receptor.";
RL      J. Neurochem. 71:1424-1434(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RA      BOGDANOV Y.D., WILDMAN S., KING B.F., BURNTOCK G.;
RL      Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; Y14705; CAA75007.1; -.
DR      EMBL; Y11433; CAA72241.1; -.
PFAM;   PF00001; 7tm_1; 1.
KW      G-protein coupled receptor.
SQ      SEQUENCE   361 AA;  40893 MW;  D1882F0C CRC32;

Query Match         81.9%;  Score 1593;  DB 11;  Length 361;
Best Local Similarity   82.7%;  Pred. No. 2e-131;
Matches   301;  Conservative   20;  Mismatches   39;  Indels   4;  Gaps   1;

Qy      1 MASTESSLLRSLGLSPGPGSSEVELDCWFDEDEKFILLPVSYAVVEVLGLGLNAPTLWLF 60
          :||||  |:|| | |||||||    | | |:||||||||||||| ||||||||||||||
Db      1 MTSAESLLFISLGPSPSSGDG----DCRFNEEFKFILLPMSYAVVFVLGLALNAPTLWLF 56

Qy     61 IFRLRPWDATATYMFHLALSDTLYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLY 120
          :|||||||||||||||||||||||||||||||| || ||||||||||||||||||||||
Db     57 LFRLRPWDATATYMFHLALSDTLYVLSLPTLVYYYAARNHWPFGTGLCKFVRFLFYWNLY 116

Qy    121 CSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVWILVVAGCLVPNLFFVTTSNKG 180
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||| |
Db    117 CSVLFLTCISVHRYLGICHPLRAIRWGRPRFAASLLCIGVWILVVAGCLVPNLFFVTTNANG 176

Qy    181 TTVLCHDTTRPEEFDHVHFSSAVMGLLFGVPCLVTLVCYGLMARRIYQPLPGSAQSSSR 240
```

Db 177 TTILCHDTTLPEEFDHVYFSSAVMVLLFGLPFLIIILVCYGLMARRLYRPLPGAGQSSSR 236

Qy 241 LRSLRTIAVVLTVFAVCFVPFHITRTIIYYLARLLEADCRVLNIVNVVYKVTRPLASANSC 300
Db 237 LRSLRTIAVVLTVFAVCFVPFHITRTIIYYQARLLQADCHVLNIVNVVYKVTRPLASANSC 296

Qy 301 LDPVLYLLTGDKYRQLRQLCGGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCSTP 360
Db 297 LDPVLYLFTGDKYRNQLQOLCRGSKFKPRTAASSLALVTLHEESISRWADTHQDSTFSAY 356

Qy 361 RADR 364
Db 357 EGDR 360

RESULT 2
ID   O57466   PRELIMINARY;   PRT;   374 AA.
AC   O57466;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   G PROTEIN COUPLED P2Y NUCLEOTIDE RECEPTOR.
GN   TP2Y.
OS   Meleagris gallopavo (Common turkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Meleagrididae; Meleagris.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 9808419.
RA   BOYER J. L., WALDO G. L., HARDEN T. K.;
RT   "Molecular cloning and expression of an avian G protein-coupled P2Y
RT   receptor.";
RL   Mol. Pharmacol. 52:928-934(1997).
DR   EMBL; AF031897; AAC60339.1; -.
DR   PFAM; PF00001; 7tm_1; 1.
SQ   SEQUENCE   374 AA; 42594 MW; 5320428C CRC32;

Query Match   57.9%; Score 1127.5; DB 13; Length 374;
Best Local Similarity   59.3%; Pred. No. 7.8e-91;
Matches 208; Conservative 56; Mismatches 70; Indels 17; Gaps 4;

Qy   9 LRSLGLSP------------GPGSSEVELDCWFDEDFKFILLPVSYAVFVLGLGLNAPTLW   58
Db   5 VRMFSLAPWTPTPTPWLGGNTIAAAEAKCVFNEEFKFILLPISYGIGTVFVVGLPLNSWAMW   64

Qy   59 LFIFRLRPWDATATYMFHLALSDTLLYVLSLPTLIYYYAAHNHWPFGTEICKFVRFLFYWN 118
Db   65 IFVSRMRPWNATTTYMFNLAISDTLYFSLPTLVYYADRNNWPFGKVFCKIVRFLFYAN 124

Qy 119 LYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAWNLVVAGCLVPNLFVTTSN 178
Db 125 LYSSLFLTCISVHRYMGICHPIRSLKWVKTKHARLICVGVWLVVTICLIPNLFVTISS 184

Qy 179 KGTTVLCHDTTRPEEFDHVHFSAAVMGLLFGVPCLVTLVCYGLMARRLYQ--PLPGSA 235
Db 185 KDNSTLCHDTTKPEEFDHVYHSSSIMALLFGIPFLVVCYCLMAKRLCKRSFPSSPSPR 244

Qy 236 QSSSRLRSTIAVVLTVFAVCFVPFHITRTIIYYLARLLEADCRVLNIVNVVYKVTRPLA 295
Db 245 VPSYKVRSIKMIILVLTVFAICFVPFHIRTLXYTSRYFQADCQTLNIINFTYKITRPLA 304

Qy 296 SANSCLDPVLYLLTGDKYRQLRQLCGGGKPQPRTAASS-LALVSLPEDSS 345
Db 305 SINSCLDPILYFMAGDKYRGRLRR---GAAQRPRVPTSLLALVSPSVDSS 352

RESULT 3
ID   O93361   PRELIMINARY;   PRT;   328 AA.

AC   O93361;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   G PROTEIN-COUPLED RECEPTOR P2Y3.
GN   P2Y3.
OS   Meleagris gallopavo (Common turkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Meleagrididae; Meleagris.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LI Q., OLESKY M., PALMER R.K., HARDEN T.K., NICHOLAS R.A.;
RT   "Evidence that the p2y3 receptor is the avian homologue of the
RT   mammalian P2Y6 receptor.";
RL   Mol. Pharmacol. 0:0-0(1998).
DR   EMBL; AF069555; AAC23863.1; -.
DR   PFAM; PF00001; 7tm_1; 1.
SQ   SEQUENCE   328 AA; 37594 MW; 1E0F2067 CRC32;

Query Match   30.8%; Score 599; DB 13; Length 328;
Best Local Similarity   42.3%; Pred. No. 7.9e-45;
Matches 127; Conservative 44; Mismatches 121; Indels 8; Gaps 4;

Qy   27 CWFDEDFKFILLPVSYAVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLLYVL   86
Db   13 CTFQEEFKQVLLPLVIYSVFLLGLPLNAAVLGOIWLARKALTRTTIIWNLATADLLIVC   72

Qy   87 SLPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRW 146
Db   73 SLPTLIYNYTQKDYWPFGGDFICKFVRRFQFYINLHGSILFLTCISVQRXYMGICHPLAS--W 130

Qy 147 GR---PRLAGLLCLAWNLVVAGCLVPNLFVTTSNKGTTVLCHDTTRPEEFDHVYHFSSA 203
Db 131 HKKKGKKLTWLVCAAVWFIVIAQCLPFFVFASTGTQRNRTVCYDLSPPDRSASYFPYGIT 190

Qy 204 VMGLLFGVPCLVTLVCYGLMARRLYQ--PLPGSAQSSSRLRSIRLRTIAVVLTVFAVCFVPF 261
Db 191 LIITGFLLPFRAILACYCSMARILCQORDELIGLAVHKKKDRAVRMIIIVVVIVFSISFPF 250

Qy 262 HITRTIIYYLARLLEA-DCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTGDKYRQLRQL 320
Db 251 HLTKTIYLIIVRSSPTLPCPTLQAFAIAYKCTRPFASMNSVLDPILFYFTQRKFRESTRYL 310

RESULT 4
ID   O15132   PRELIMINARY;   PRT;   370 AA.
AC   O15132;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   P2Y5-LIKE RECEPTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97366605.
RA   JANSSENS R., BOEYNAEMS J.M., GODART M., COMMUNI D.;
RT   "Cloning of a human heptahelical receptor closely related to the P2Y5
RT   receptor.";
RL   Biochem. Biophys. Res. Commun. 236:106-112(1997).
DR   EMBL; AF005419; AAB66322.1; -.
DR   PFAM; PF00001; 7tm_1; 1.
SQ   SEQUENCE   370 AA; 41861 MW; A42AB01D CRC32;

Query Match   23.2%; Score 450.5; DB 4; Length 370;
Best Local Similarity   33.4%; Pred. No. 7.9e-32;
Matches 102; Conservative 67; Mismatches 131; Indels 5; Gaps 5;

Qy   14 LSPGPGSSEVELDCWFDEDFKFILLPVSYAVFVLGLGLNAPTLWLFIFRLRPWDATATY   73

```
Db     18  LRPRLGNATANNTCIVDDSFKINLNGAVYSVVFILGLIINSVSLFVCFRMKRSETAIF  77
               |:| |    | |  |: |  |  |||| : :|||  || ||||| || |:| |
Qy         MFHIALSDTLYVLSLPFLIYIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHR  77

Db     74  VWRTYLSKITIFIEVGFIIPLLLNVSCSSVVLRTLRKPATLSQIGTNKKKVLKMIIVHM  255
Qy    196  VWRTYLSKITIFIEVGFIIPLLLNVSCSSVVLRTLRKPATLSQIGTNKKKVLKMIIVHM  255

Db     78  ITNIAVSDLLFVCTLPFKI-FYNFNRHWPFGDILCKISGTAFLTNIIGSMLFLTCISVDR  136
              :|||||||    |  |     |   |  ||:|||  :|| |   |  |  |  |   |
Qy    252  TVFAVCFVPFHITRTIYYLARLLE-ADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTG  310

Db    134  YLGICHPLRALRWGRPRLAGLLCLAVV-LVVAGCLVPNLFFVTTSNKGTTVLCHDTTPE  192
              :||||| ||||:||||||| ||| |  |:|| || |  |||  | |  |   |
Qy    137  FLAIYPFPFRSRFIRTRRNSAIVCAGVWILVLSLGGGISASLFSTTNVNNATTT-CFEGLSKR  195

Db    134  YLGICHPLRALRWGRPRLAGLLCLAVV  159
```

```
RESULT  5

ID    P79960         PRELIMINARY;    PRT;   353 AA.
AC    P79960;
DT    01-MAY-1997 (TrEMBLrel. 03, Created)
DT    01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT    01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE    MESENCHYME-ASSOCIATED SERPENTINE RECEPTOR.
OS    X-MSR.
OS    Xenopus laevis (African clawed frog).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC    Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC    Xenopus.
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE; 97109511.
RA    DEVIC E., PAQUEREAU L., VERNIER P., KNIBIEHLER B., AUDIGIER Y.;
RL    "Expression of a new G protein-coupled receptor X-msr is associated
RL    with an endothelial lineage in Xenopus laevis.";
RL    Mech. Dev. 59:129-140(1996).
CC    -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC    -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR    EMBL; X93045; CAA63612.1; -.
DR    PFAM; PF00001; 7tm_1; 1.
DR    PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW    G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ    SEQUENCE  353 AA;  40303 MW;  3354C3AE CRC32;
```

```
Query Match            20.6%;  Score 400;  DB 13;  Length 353;
Best Local Similarity  33.7%;  Pred. No. 1.9e-27;
Matches 106;  Conservative  59;  Mismatches 130;  Indels  20;  Gaps  10;

Qy     30  DEDEKFILLPVSYAVVFVIGLGLNAPTLWLFIFLRLRPWDATA-TYMFHLALSDTLYVLSL  88
               |  ||:  :||   | ||| ||| |   || | |  | :  :   ||||||||||
Db     30  DWDFSYSLLPVFYMIVFVLSGNGVVIFT-VWKSKPKRRSADTYIGNIALADIAFVVTL  88

Qy     89  PTLIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGR  148
              |  |||:  | || |   | | | | | |   | || |||:||||| ||||||| ||
Db     89  PLWARTYALGFHWPGSALCKLSSYLVLLNMFASVFCTCLSFDRVLAIVHSLSSAKL-R  147

Qy    149  PRLAGLLCLAV-WLVVAGCLVPNLFFVTTSNKGTTVLC----HDTTRPEEFDHY----VHF  200
              |||||  ||  ||  || |||||  | : :     |       || | :||       |::
Db    148  SRPSIIVSLAVIWLFSGLLALPSLLLRDTRVEGNNTICDLPFSGVSSKENENFWIGLSI  207

Qy    201  SSAVMGLLRFGVPCLVTLVCYGLMARRL---YQPLPGSAQSSSRLRSLRTIAVVLTVFAVC  257
              ::| :|||  ||| | | ||| |  |       | | |  |  :| |||| |  | |
Db    208  LTTVPGFL--LPLLMTIFYCPIGGKVTMHFQNLKEEQKKRRL--LKIITILVVVFAIC  263
```

```
RESULT  6

ID    O15185         PRELIMINARY;    PRT;   415 AA.
AC    O15185;
DT    01-JAN-1998 (TrEMBLrel. 05, Created)
DT    01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT    01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE    G PROTEIN-COUPLED RECEPTOR CKR-L2.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC    Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN    [1]
RP    SEQUENCE FROM N.A.
RA    GUTIERREZ J., VARONA R., ZABALLOS A., LIND P., MARQUEZ G.;
RL    Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; Z79783; CAB02143.1; -.
DR    PFAM; PF00001; 7tm_1; 1.
SQ    SEQUENCE  415 AA;  45608 MW;  E80DF045 CRC32;
```

```
Qy    258  FVPFHITRTIYYLARL--LEADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTGDKYRR  315
              :|||||   |  || |   |: || | |  |  :||||| || |||||||||||||||
Db    264  WLPFHILKTIHFLDLMGFLELSCSTQNIIVSLHPYATCLAYVNSCLNPFLIYAFFDLRFRS  323

Qy    316  QLRQLCGGGKPQPRI  330
              || ||| |||  | |
Db    324  QCFFFWFQKSPSPRI  338
```

```
Query Match            19.2%;  Score 374;  DB 4;  Length 415;
Best Local Similarity  30.6%;  Pred. No. 4.1e-25;
Matches 114;  Conservative  57;  Mismatches 155;  Indels  46;  Gaps  11;

Qy      6  SSLLRSLGLSPGPGSSEVELDCW------FDEDEKFILLPVSYAVVFVIGLGLNAPTLW  58
              ::|:|  |  |  |  : ||       | |  ||  ||||  ||   |  ||| |
Db     64  AALLENFSSSYDYGENESDSCCTSPPCPQDFSLNFDRAFLPALYSLLFLGLLGNGAVRA  123

Qy     59  LFIFLRLRPWDATATYMFHLALSDTLYVLSLPFLIYIYYAAHNHWPFGTEICKFVRFLFWN  118
              | |  || |  :|| |  |||||||||  ||  | | | |    ||| || |  ||| |
Db    124  VLLSRRIALSSTDTFLLHLAVADTLLVLTLP--LWAVDAAVQWVFGSGLCKVAGALFNIN  181

Qy    119  LYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVV--VPCLVPNLFFVTTSN  178
              |||||| | ||:||| ||||||||||||| || |  | | |  :||| || | |   |
Db    182  FYAGALLLACISFDRYLNIVHATQLYRGRGPPARVTLTCLAVWGICLFALPDFIFLSAH-  241

Qy    179  KGTTVLCHD-----TTRPEEFDHYVHFSSAVMGLLFG--VPCLVTLVCYG-LMARRLYQP  230
              | |  | ||       | || |  | |   :|| |     ||| | |   | |||| | |
Db    241  -----HDERLNATHCQYNFPQVGRIALRVLQLVAGFLLPLVMAYCYAHILAVLV--292

Qy    231  LPGSAQSSSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYYLARL--LEADCRVLNIVNVVY  288
              | :| |  :| |||| || | |    | |||||   |  |  || |   |: || | |  |
Db    292  ----SRGQRLRAMRLVVVVVAFALCWTPYHLVHLVLVDITIMDLGALAARNCGRESRVDVAK  347
```

```
RESULT  7

ID    O42324         PRELIMINARY;    PRT;   383 AA.
AC    O42324;
DT    01-JAN-1998 (TrEMBLrel. 05, Created)
DT    01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT    01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE    MU-OPIOID RECEPTOR.
OS    Catostomus commersoni (white sucker).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC    Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
```

```
Qy    344  SSCRWAATPQDS  355
              || :   ||
Db    402  SS--WSEFSEAS  411

Qy    289  KVTRPLASANSCLDPVLYLLTGDKYRRQ-----LRQLCGGGKPQPRTAASSLALVSLPED  343
              : |||| || ||||| |  |  : ||      |||| ||||  |  | :|   |  ||
Db    348  SVTSFGIGYMHCCLNPLIYAFVGVKFRERMWMLLLRLCGPNQRGLQRQPSSS-----RRD  401
```

OC Cobitoidea; Catostomidae; Catostomus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CNS;
RA DARLISON M.G., HARVEY R.J., GRETEN F.R., KREIENKAMP H.J., ZWIERS H.,
RA STHMER T., LEDERIS K., RICHTER D.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; Y10904; CAA71843.1; -.
DR PFAM; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ SEQUENCE 383 AA; 43232 MW; 00BBDCBD CRC32;

Query Match   18.9%; Score 368; DB 13; Length 383;
Best Local Similarity   29.7%; Pred. No. 1.3e-24;
Matches 105; Conservative 63; Mismatches 128; Indels 58; Gaps 13;

Qy   36 ILLPVSYAVVFVLGLGLNAPTLMLFIFRLRPWDATATYMFHLALSDTLYVLSLPTLIYY   95
         ||| ||:|||:| :||| ||||:||:||||||:|||:|||| |||||||||||||||||
Db   58 IIITLLYSIVCVGLVGNVLVMYVIIRYTKMKTATNIYIFNLALADALATSTLPFQSVNY 117

Qy   96 AAHNHWPFGETEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLL 155
         |||:|||||||||||||||||||||||:|| |:|||||||||:|||||||| ||| ||
Db 118 LM-GTWPFGDVVCKIVMSIDYNFTSIFTLLTTMSIDRYIAVCHPVKALDFRTPRNAKIV 176

Qy 156 CLAVWLLVVAGCLVPNLFVTT----------SNKGTTVLCHDTTRPEEFDHIVVHF---- 201
         |||:||:|||| :|| |||| ||||||||||:||:|:
Db 177 NVCNWILSSAGLPVWVMASTTIENONSPLQVSNFDCTLL----------:FPHPPWWETL 227

Qy 201 -SSAVMGLLFGVPCLVTLVCYGLMARRL--YQPLPGGSAQSSSRLRSL-RTIAVVLTVFAV 256
         :| || :|| |||:|||||||:|||| |:|:| ||| ::|:|:| ||:|| ||:||
Db 228 LKICVFILAFIMPVLLITVCVGLMILRLKSVRMLSGGSKEKDRNLRRITRMVLVVVAVFII 287

Qy 257 CFVPFHITRTIYYLARLLEADCRVLNIVINVNVYKVTR------PLASANSCLDPVLLLTGD 311
         |||||:||| ||||||||||||||||||:|||:||||| ||:||||||||||||||
Db 288 CWTPIH----IFVIIKAL------VIIPNSLFQTVTWHFCIALGYTNSCLNPVLYAFLDE 337

Qy 312 KYRRQLRQLCGGGGKPQPRTAASSLALVSLPEDSSCRWAATPQ-DSSCSTPRADR 364
         |||:|||:| | :|||:| | ::|||| |: || |
Db 338 NFKRCFRFC---VPSP---SVLDLQNSTRNSN----------POCEGQSSGHKVDR 378

───

RESULT 8
ID O77590
AC O77590;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT 01-AUG-1999 (TrEMBLrel. 11, Last annotation update)
DE ANGIOTENSIN II TYPE-1 RECEPTOR.
GN AT1.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RA MILLICAN D.S., BIRD I.M.;
RT "Isolation of an ovine genomic sequence containing the full-length
RT angiotensin II type-1 receptor.";
RL Endocr. Res. 0:0-0(1999).
RN [2]
RP SEQUENCE OF 1-132 FROM N.A.
RA MILLICAN D.S., MAGNESS R.R.;
RT TISSUE=ADRENAL CORTEX;
RX MEDLINE; 98351904.
RT BIRD I.M., MILLICAN D.S., MAGNESS R.R.;
RT "Specific pregnancy-induced angiotensin II type-1 receptor expression
RT in ovine uterine artery does not involve formation of alternate
RT splice variants or alternate promoter usage.";
RL Biol. Reprod. 59:219-224(1998).

───

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AF069750; AAC99344.1; -.
DR EMBL; AF056308; AAC32613.1; -.
KW Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ SEQUENCE 359 AA; 41046 MW; AD06F071 CRC32;

Query Match   18.3%; Score 356.5; DB 6; Length 359;
Best Local Similarity   27.7%; Pred. No. 1.2e-23;
Matches 90; Conservative 66; Mismatches 150; Indels 19; Gaps 5;

Qy   2 ASTESSLLRSLGLSPGGSSEVELDCWEDEDKFILLPVSYAVVFVLGLGLNAPTLWLFI   61
         :||| | |:: | | :   ::||||:::::|||::|:|:|:|||| :|| |
Db   5 SSTEDGIKRIQDDCPKAGR----------HNYIFIMIPTLYSIIFVVGLFGNSLVVIV   54

Qy   62 FRLRPWDATATYMFHLALSDTLYVLSLPTLIYYAAHNHWPFGETEICKFVRFLFYWNLYC 121
         | ||| |||||:| ||||| |||:||:|:|||| |||||||| |||||||||| ||
Db   55 FYMKLKTVASVFLLNLALADLCFLLTLPLWAVYTAMEYRWPFGNYLCKIASGSVSFNLYA 114

Qy 122 SVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVWLVVAGCLVP-----NLFVTT 176
         |||| |:||:| : | ||:|||| |||| || | |||: || || |||||
Db 115 SVFLLTCLSIDRYLAIVHPMKSRLRRTMLVAKYTCIIIWLLAGLLASLPTIIHRNVFFIEN 174

Qy 177 SNKGTTVLCHDTTRPEEFDHIVVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQ 236
         | :| ||||||||||||||:|:||:   :| |: |||:|| |||:||| |:|:|
Db 175 TNI-TVCAFHYESQNSTLPVGLGLITKNIIGFLF--PFLILITSYTLIWKTLKKAYEIQKN 231

Qy 237 SSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYYLARL--LEADCRVLNIVNVVKVTRPLA 295
         | :|||:| |: ||:| | ||:||||||:||||||| |||| |:|| |:||| |
Db 232 KPRKDDIFKILLAIVLFFFFSWVPHQIFTFMDVLIQLGLIRDCKIEDIVDTAMPITICLA 291

Qy 296 SANSCLDPVLLLTGDKYRRQLRQL 320
         | ||| |||||||| |
Db 292 YFNNCLNPFFYGPLGKKFKYFLQL 316

───

RESULT 9
ID O00398
AC O00398;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PUTATIVE PURINERGIC RECEPTOR P2Y10.
GN DJ333E23.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA BOHM S.K.;
RL Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA GRAFHAM D.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF000545; AAB57836.1; -.
DR EMBL; Z82200; CAB36889.1; -.
PFAM; PF00001; 7tm_1; 1.
KW Receptor; Transmembrane.
SQ SEQUENCE 339 AA; 38774 MW; AED97BCE CRC32;

Query Match   18.2%; Score 354; DB 4; Length 339;
Best Local Similarity   29.7%; Pred. No. 1.9e-23;
Matches 88; Conservative 55; Mismatches 131; Indels 22; Gaps 7;

Qy   33 FKFILLPVSYAVVFVLGLGLNAPTLMLFIFRLRPWDATATYMFHLALSDTLYVLSPTLI   92
         | :|||||| ||:||:|| ::| | :|| | ||||:|||:|||||||| :|||||| |
Db   32 FQYSLVATYILIFIPGLLANSAALWVLCRFISKKNKAIIFMINLSVADLAHVLSLPLRI   91

Qy   93 YYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLA 152

```
Db    92  YYIYSH-HWPFQRALCLLCFYLKYLNMYASICFLLTCISLQRCFFLLKPFRARQWKRRYDV  150
           ||:||  ||:|| ||:||| ||:| :| ||| :  ||| || ||| | | :|||| :|
Qy   153  GLLCLAVWLVVAGCLVPNLFYVTL----SNKGTTVLCHDTTRPEEFDHYVHFSSAVMGLL-   209
           | |||||   || :|| ||::  |       |||||:| |||:|||: | || :||| :
Db   151  G-ISAAIWIVGTACLPPPILRSTDLNNKS----CFADLGYKQMN----AVALVGMIT      200

Qy   209  ------FGVPCLYTLVCYGLMARRLYQPLPGSAQSSRLRSLRTIAVVLTVFAVCFVPFH    262
                 ||||| | :|| |::   || |||  ||||| || :| ||| || | ||||
Db   209  VAELAGFVIPVIIIAWCTWKTTISLRQPPMAFQGISEROKALRVVFMCQAAVFFICFTPYH  260

Qy   263  ITRTIYYLAR-LLEADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTGDKYRRQL    317
           |||:||| |  |||||:|| |:|      | ||||  ||||| || || || |
Db   261  INFIFYTMVKEIIISSCPVVRIALYFHPFCLCLASLCCLLDPILYYFMASEFRDQL    316

RESULT  10
ID      P70058             PRELIMINARY;        PRT;     303 AA.
AC      P70058;
DT      01-FEB-1997 (TrEMBLrel. 02, Created)
DT      01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT      01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE      ANGIOTENSIN RECEPTOR RELATED PROTEIN.
GN      XANGIO1.
OS      Xenopus laevis (African clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC      Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC      Xenopus.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      SAHA M.S., OAKES J.A., MILES R.R.;
RL      Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC      -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC      -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR      EMBL; U72029; AAB17004.1; -.
DR      PFAM; PF00001; 7tm_1; 1.
DR      PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW      G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ      SEQUENCE  303 AA;  34088 MW;  F04BBF0B CRC32;

Query Match        18.1%;   Score 352;  DB 13;  Length 303;
Best Local Similarity  33.8%;   Pred. No. 2.5e-23;
Matches  91;  Conservative  54;  Mismatches 104;  Indels  20;  Gaps  10;

Qy    30  DEDEKFILLPVSYAVVFVLGLGLNAPTLWLFIFRLRPWDAFA-TYMFHLALSDTLYVLSL    88
           | ||| ||  |: |  |  ||||  :|| ||:|  | :: ||:||| || || |
Db    30  DWDFSYSLLPVFYIMIVFVLGLSGNGVVIFT-VWKSKPRRSADTYIGNLALADLAEVVTL   88

Qy    89  PTLIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGR   148
           | |||:|  :|||| | || ||| | ||||| | ||:||||  || ||||||||| ||
Db    89  PLWAYTALGFHWPFGSSALCKLSSSIVLLNMFASVFCLTCLSFPDRILAIVHSLSSAKL-R  147

Qy   149  PRLAGLLCLAV-WLVVAGCLVPNLFFVTTISNKGTTIVLC---HDTTRPEEFDHY----VHF  200
           ||| |||| |  ||||||||||||| | ::     |||||:|||:|| :| ||||||  : |
Db   148  SRSSIVSLAVIWLFSGLLALPSLILRDTRVEGNNTICDLDFSGVSSKENENFWIGGLSI  207

Qy   201  SSAVMGLLFGVPCLYTLVCYGLMARRL---YQPLPGSAQSSRLRSLRTIAVVLTVFAVC   257
           | |||  |  |  | : ::  |::        || ||||  |||| || || || ||  |
Db   208  LTTTVPGFL--LPLLMTIFYCFIGGKVTMHFQNLKEEQKKKRL--LKIIILVVVFAIC   263

Qy   258  FVPFHTRTIYYLARL--LEADCRVLNIV    284
           : :||  |||| ||   |||||:|| ||
Db   264  WLPFHILKTIHFLDLMGFLELSCSTONII  292

RESULT  11
ID      O57585             PRELIMINARY;        PRT;     373 AA.
AC      O57585;
DT      01-JUN-1998 (TrEMBLrel. 06, Created)
DT      01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
```

```
DT      01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE      OPIOID RECEPTOR HOMOLOGUE.
GN      OPRD1.
OS      Brachydanio rerio (zebrafish) (zebra danio).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC      Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC      Cyprinoidea; Cyprinidae; Rasborinae; Danio.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      BARRALLO A., GONZALEZ-SARMIENTO R., GARCIA-MALVAR F., RODRIGUEZ R.E.;
RL      Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC      -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC      -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR      EMBL; AJ001596; CAA04862.1; -.
DR      PFAM; PF00001; 7tm_1; 1.
DR      PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR      ZFIN; ZDB-GENE-990415-199; Oprd1.
KW      G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ      SEQUENCE  373 AA;  42520 MW;  E900D477 CRC32;

Query Match        18.1%;   Score 352;  DB 13;  Length 373;
Best Local Similarity  31.1%;   Pred. No. 3.1e-23;
Matches  95;  Conservative  57;  Mismatches 117;  Indels  36;  Gaps  10;

Qy    36  ILLPVS----YAVVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPTL    91
           ||| :     | : |: ||||  :|| || :|:   |: |||| || || ||||| ||
Db    49  IIIAISITALYSVICVVGLLGNLLVMYGVVRYTKLKTATNIYIFNLALADALATSTLPFQ  108

Qy    92  IYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRL   151
           || ||| ||||| | ||  | || || :| ||| | |||  ||||||||||||| | |
Db   109  STKYLM-NTWPFGELLCKVVIAIDYYNMFTSIFTLTMMSVDRYIAVCHPVRALEFRTPIK  167

Qy   152  AGLLCLAVWLVVAGCLVPNLFFVTT--SNKGTTVLCHDTTRPEEFDHYVHFSSAVMGLLF  209
           ||:|||: ||||||||||||||| |   |||||:| ||:||:||:| || || ||:|| |
Db   152  AKIINVCIWILSSAVGVPIMIMAVTRVTNQNTTVCMLKFPDP---DWYWDTVRKICVIFF  224

Qy   210  G--VPCLYTLVCYGLMARRL--YQPLPGSAQSSRLRSL-RTIAVVLTVFAVCFVPFHIT  264
            |  |||| | :|| :::    :|| ||||  ||||:|: || || || |||| || |
Db   225  AFVVPVLVTICYGLMILRLKSVRLLSGGSKEKDRNMRRITRMVLVVVAAFICWTPIHIF  284

Qy   265  RTIYYLARLLEADCRVLNIVNVVYKVTRPLAS-------ANSCLDPVLYLLTGDKYRRQ  316
           |||||  |||||| || |: |  : ||| ||||       | || |||||| |||| |
Db   285  -------IIEK----TLVDINQKNPFVIASWHLHRTGYTNSSLNPVLYAFLDENFKRC  331

Qy   317  LRQLC  321
           || ||
Db   332  FRDFC  336

RESULT  12
ID      O89609             PRELIMINARY;        PRT;     383 AA.
AC      O89609;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT      01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE      G PROTEIN-COUPLED RECEPTOR.
OS      Equine herpesvirus 2.
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Betaherpesvirinae.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=86/67;
RX      MEDLINE; 95302501.
RA      TELFORD E.A., WATSON M.S., AIRD H.C., PERRY J., DAVISON A.J.;
RT      "The DNA sequence of equine herpesvirus 2.";
RL      J. Mol. Biol. 249:520-528(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=86/67;
RA      TELFORD E.A.R.;
RL      Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; U20824; AAC13866.1; -.
DR   EMBL; U20824; AAC13788.1; -.
DR   PFAM; PF00001; 7tm_1; -.
SQ   SEQUENCE   383 AA;  43667 MW;  60F5BFD4 CRC32;

Query Match             18.0%;  Score 351;  DB 12;  Length 383;
Best Local Similarity   26.3%;  Pred. No. 3.8e-23;
Matches   97;  Conservative   74;  Mismatches  150;  Indels   48;  Gaps    9;

Qy     1  MASTESSLLRSL-----GLSPGPGSSEVELDCWFDEDFK------------FILLPVS   41
                :||:|||    :   :   ::     ::                   :::
Db    19  MTTNFTSLITSVVTTIASLVPSTNSSEDYDDLDDVDYEESAPCYKSDTTRLAAQVVPAL   78

Qy    42  YAVAVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDLLYVLSLPTLIYYAAHNHW  101
          :  :  | || || | |:: |  |  ::    ::  |||:: ||||::  |   :    |:
Db    79  YLIVFLFGLLGNILVIVIIVIRYMKIKNLTNMLLLNLAISDLLFLITLPFWMHYIGMYHDW  138

Qy   102  PFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAWL  161
          :||: :|||:||:|    | :|| ::: |  ::| |   ::     :::|:|::   :|
Db   139  TFGISLCKLLRGVCTMSLYSQVFCIILLTVDRYLAVVYAVTALRERTVTCGIVTCVCTWF  198

Qy   162  VVAGCLVPNLFFVTTSNKGTTVLVLCHDTTRPEEFDHY---VHFSSAVMGLLFGVPCLVTLV  218
          :  ||::   :| : : :   |:|:: |::       |    :::  :| || :: |:|:|
Db   199  LAGLLSLPEFFHGHQDDNGRVQC-DPYYPEMSTNVWRAHVAKVIM-LSLILPLLIMAV  256

Qy   219  CYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHIRTI-YYLARLLEAD  277
          || : ::||   :        ::  ::   : | : | :|| :   : : ||||||
Db   257  CYYYIRRLLR-----RPSKKKYKAIRLIFVIMVAYFVFWTPYNIVLLSTFHATLLNLQ  311

Qy   278  CRVLNIVNVVYKVTRPLASANSCLDPVLLLTGDKYRRQLRQ----------LC------  322
          |  :       : :| | ||||||||||:: |||::|::|           :|
Db   312  CALSSNLDMALLITKIVAYTHCCINPVIYAPVGGEKFRRHLYHFFHYTVAIYLCKYIPFLS  371

Qy   322  --GGKPQP  328
            ||  ||
Db   372  GDGEGKEGP  380
```

```
RESULT 13
O18951

ID   O18951     PRELIMINARY;       PRT;   133 AA.
AC   O18951;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   P2U/P2Y2 RECEPTOR (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   CHANG A.S., CHANG S.M., SCHILLING W.P.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF005153; AAB62543.1; -.
DR   PFAM; PF00001; 7tm_1; -.
FT   NON_TER         1      1.
FT   NON_TER       133    133.
SQ   SEQUENCE   133 AA;  15006 MW;  190A8B1F CRC32;

Query Match             17.9%;  Score 349;  DB 6;  Length 133;
Best Local Similarity   51.1%;  Pred. No. 1.9e-23;
Matches   68;  Conservative   23;  Mismatches   40;  Indels    2;  Gaps    1;

Qy   129  ISVHRYLGICHPLRALRWGRPRLAGLLCLAWLVVAGCLVPNLFFVTTSNKGTTVLVLCHDT  188
          ||||| |||  |:|| ||:|:| | ||  |  ::  ||||:|||:||||| : :::||||
Db     1  ISVHRCLGIVRPLSLRWGRARYARRVAFAVVWVLYLXCOAPVLFYVTTSRSSRICHDT   60

Qy   189  TRPEEFDHYVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRL--RSLRT  246
          :  |:| :|:|:::::| :|:: |  ::|||| |:|||||        ::|:   :||:|
Db    61  SARELFSHEVAYSSVMLSLLFAAPFAVILVCYVLMARRLLKPAYGTSGGLPRAKRKSVRI  120
```

```
Qy   247  IAVLTVFAVCFV  259
          ||::||| :||:
Db   121  IAIVLTVFVLCFL  133
```

```
RESULT 14
O95853

ID   O95853     PRELIMINARY;       PRT;   381 AA.
AC   O95853;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ORPHAN G PROTEIN-COUPLED RECEPTOR.
GN   GPR34.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99156852.
RA   MARCHESE A., SAWZDARGO M., NGUYEN T., CHENG R., HENG H.H., NOWAK T.,
RA   IM D.S., LYNCH K.R., GEORGE S.R., O'DOWD B.F.;
RT   "Discovery of three novel orphan G-protein-coupled receptors.";
RL   Genomics 56:12-21(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; AF118670; AAD17248.1; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ   SEQUENCE   381 AA;  43860 MW;  07C967ED CRC32;

Query Match             17.9%;  Score 348.5;  DB 4;  Length 381;
Best Local Similarity   27.1%;  Pred. No. 6.3e-23;
Matches   95;  Conservative   68;  Mismatches  156;  Indels   31;  Gaps    6;

Qy    27  CWFDEDFKFILLPVSYAVAVFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDLLYVL   86
          |  |: :|:  : :| :: : |:|| ||  : :  |     :  ::: :::: |||:   :
Db    46  CPMDEKLLSTVLTSVSVIFIVGLVGNIIALIVFLGIHRKRNSIQIYLLNVAIADLLLIF  105

Qy    87  SLPTLIYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRW  146
          :|| :::   :: |  :| :||:: :||  : | :: ::  ::  ::| |    :  : :
Db   106  CLPFRIMYHINQNKWTLGVILCKVVGTLFYMNMYISIILLGFISLDRYIKINRSIQQRKA  165

Qy   147  GRPRLAGLLCLAWLVVAGCLVPNLFFVTTSNKGTTVLVLCHDTTRPEEFDHYVHFSSA---  204
          :   : |:  : | : |  : | :::: :: : :  : : :           :|:::::
Db   166  ITTKQSIYVCCIVWMLALGGFLTMIILTLKKGGHNSTMCF------------HYRDKHNAKGE  216

Qy   204  ----VMGLLFGVPCLVTLVCYGLMARRLYQ-------PLPGSAQSSSRLRSLRTIAVVLT  252
              ::|::  : :::::: :: | :  ::          : ||  :::  :  ::  :::
Db   217  AIFNFILVVMFWLIFLLIILSYIKIGKNLRLRISKRRSKFPNSGKYAT---TARNSFIVLI  273

Qy   253  VFAVCFVPFHITRTIYYLARLLEADCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTGDK  312
          :| :||||:|: | |:: ::: :    ::  :  :  ::: | :: :|||||: : : ::
Db   274  IFTICFVPYHAFRFIYISSQLNVSSCYWKEIVHKINEIMLVLSSFNSCLDPVMYFLMSSN  333

Qy   313  YRRQLRQLC---GGGKPQPRTAASSLALVSLPEDSSCRWAATPQDSSCST  359
          :|::: ||    :   :                 :::     ::::|:||
Db   334  IRKIMCQLLFRRFQGEPSRSESTSEFKPGYSLHDTSV--AVKIQSSSKST  381
```

```
RESULT 15
O43190

ID   O43190     PRELIMINARY;       PRT;   371 AA.
AC   O43190;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PURINERGIC P2Y11 RECEPTOR.
GN   P2Y11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   COMMUNI D., GOVAERTS C., PARMENTIER M., BOEYNAEMS J.M.;
RL   J. Biol. Chem. 0:0-0(1997).
DR   EMBL; AF030335; AAB88674.1; -.
DR   PFAM; PF00001; 7tm_1; 1.
SQ   SEQUENCE   371 AA;  40174 MW;  5686F41C CRC32;
```

```
Query Match            17.6%;  Score 342;  DB 4;  Length 371;
Best Local Similarity   32.0%;  Pred. No. 2.3e-22;
Matches   97;  Conservative  44;  Mismatches 132;  Indels  30;  Gaps  7;

Qy    29  FDEDFKFILLPVSYAVVFVLGLGLNAPTLWLF-IFRLRPWDATATYMFHLALSDTLYVLS  87
              ||||||  ||| :| :  |  |  ||||||  ||  | |  |  ||| ||||||:|:|:
Db    23  FQGDFLWPIL---VVEFLVAVASNGLALYRFSIRKQRPWHPAVVFSVQLAVSDLLCALT  78

Qy    88  LPTLIYYYAAHNHWPFGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWG  147
              ||-| || || |||  ||  |||||:|||||: |::||| | |  |
Db    79  LPPLAAYLYPPKHWRYGEAACRLERFLFTCNLLGSVIFITCISLNRYLGIVHPFFARSHL  138

Qy   148  RPRLAGLLCLAVWLVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPE------EFDH----  197
              || :: :  |  |  :||  |||| ||  ||            ||    |||
Db   139  RPKHAWVSAAGWVLAALLAMPTLSFSHLKRPQQGAGNCSVARPEACIKCLGTPADHGLAA  198

Qy   197  YVHFSSAVMGLLFGVPCLVTLVCYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAV  256
              | :|  || |||| :||  |:|| |  | ||   ||   :|     ||   :||
Db   199  YRAYSLVLAGLGCGLPLLLTLAAYGALGRAVLRS-PGMT-VAEKLRVALVASGVALYAS  256

Qy   257  CFVPFHITRTIYYLAR------------LLEADCRVLNIVNVVYKVTRPLASANSCLDPV  304
              :||:||| || |    |         : :|| |  |  ||  | |:  |:
Db   257  SYVPYHIMRVLNVDARRWSTRCPSFADIAQATAALELGPYVGYQVMRGLMPLAFCVHPL  316

Qy   305  LYL  307
              || :
Db   317  LYM  319
```

OM nucleic - nucleic search, using sw model

GenCore version 4.5.
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: December 29, 1999, 03:43:59 ; Search time 727.51 seconds
(without alignments)
6733.258 Million cell updates/sec

Title: US-09-077-173A-1
Perfect score: 1429
Sequence: 1 AAGGGAGCTTGGGTAGGGGC..........GCCACGGAGGCAGGGATATC 1429

Scoring table: IDENTITY_NUC

Searched: 4089388 seqs, 1713965092 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold :   8178776

1: em_est1:*	50: gb_est31:*
2: em_est2:*	51: gb_est32:*
3: em_est3:*	52: em_est20:*
4: em_est4:*	53: em_est21:*
5: em_est5:*	54: em_est22:*
6: em_est6:*	55: em_est23:*
7: em_est7:*	56: em_est24:*
8: em_est8:*	57: em_est25:*
9: em_est9:*	58: em_est26:*
10: em_est10:*	59: gb_est38:*
11: em_est11:*	60: gb_est35:*
12: em_est12:*	61: gb_est36:*
13: em_est13:*	62: gb_est37:*
14: em_est14:*	63: gb_est38:*
15: em_est15:*	64: gb_est38:*
16: em_est16:*	65: em_est27:*
17: em_est17:*	66: em_est28:*
18: em_est18:*	67: em_est29:*
19: em_est19:*	68: em_est30:*
20: gb_est8:*	69: gb_gss1:*
21: gb_est2:*	70: gb_gss2:*
22: gb_est3:*	71: gb_gss3:*
23: gb_est4:*	72: gb_gss4:*
24: gb_est5:*	73: gb_gss4:*
25: gb_est6:*	74: em_gss1:*
26: gb_est7:*	75: em_gss2:*
27: gb_est8:*	76: em_gss3:*
28: gb_est9:*	77: gb_gss5:*
29: gb_est10:*	78: gb_gss6:*
30: gb_est11:*	79: gb_gss6:*
31: gb_est12:*	80: gb_gss8:*
32: gb_est13:*	81: gb_gss9:*
33: gb_est14:*	82: em_gss5:*
34: gb_est15:*	83: em_gss6:*
35: gb_est16:*	84: em_gss7:*
36: gb_est17:*	85: em_gss8:*
37: gb_est18:*	86: em_gss9:*
38: gb_est19:*	87: em_gss10:*
39: gb_est20:*	88: em_gss11:*
40: gb_est21:*	89: gb_gss10:*
41: gb_est22:*	90: gb_gss11:*
42: gb_est23:*	
43: gb_est24:*	
44: gb_est25:*	
45: gb_est26:*	
46: gb_est27:*	
47: gb_est28:*	
48: gb_est29:*	
49: gb_est30:*	

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		130.8	9.2	527	63	AI956440	AI956440 u177c02.y
2		117.2	8.2	586	69	FR0019937	AL012824 F.rubripe
3	c	80.2	5.6	544	45	AI344017	AI344017 tc01e02.x
4	c	75.4	5.3	542	63	AI942392	AI942392 wo77g12.x
5	c	74.2	5.2	466	21	R15256	R15256 yf89d03.r1
6		74.2	5.2	316	31	AA296456	AA296456 EST11038
7		73.8	5.2	261	32	AA331167	AA331167 EST35211
8		70.8	5.0	599	38	AA804531	AA804531 ns28c04.s
9		70	4.9	592	64	AL048512	AL048512 DKFZp586M
10		66	4.6	529	29	AA177828	AA177828 mt07f02.r
11		66	4.6	650	36	AA611127	AA611127 vo61f08.r
12		65.8	4.6	476	48	AI607213	AI607213 vo61f08.y
13	c	65.4	4.6	816	30	AA205847	AA205847 zq50c02.r
14		65.4	4.6	570	51	AI739295	AI739295 wl30b03.x
15		62.6	4.4	559	30	N93987	N93987 za66c09.r1
16		62.6	4.4	249	23	AA239704	AA239704 my15c07.r
17		60.4	4.2	274	21	H29103	H29103 ym31f07.r1
18	c	59.2	4.1	520	49	F03704	F03704 HSC27C022 n
19		58.2	4.1	361	21	AI663305	AI663305 uk27c10.y
20		57.2	4.0	480	21	F07588	F07588 HSC2BD071 n
21		57.2	4.0	480	23	H20663	H20663 ym47h08.r1
22		56.4	3.9	733	63	AI959485	AI959485 fd10g03.y

23	56	3.9	504	37	AA690768	AA690768 vt25e01.r
24	55	3.8	456	32	AA354393	AA354393 EST62864
25	54.2	3.8	453	25	H97311	H97311 EST42j10 WA
26	54.2	3.8	478	37	AA672048	AA672048 v107f01.r
27	53	3.7	575	63	AI979910	AI979910 v107f01.r
28	52.2	3.7	446	42	AI131863	AI131863 pat.pK000
29	51.2	3.6	434	44	AI263529	AI263529 uc36h08.r
30	50	3.5	486	31	AA290864	AA290864 g120c06.x
31	49.6	3.5	461	50	AI670734	AI670734 zt19a09.s
32	49.6	3.5	385	21	R91585	R91585 wc28e08.x
33	49.4	3.5	484	46	AI417609	AI417609 yq10a01.r1
34	49.2	3.4	970	69	CNS01C09	CNS01C09 tg80b06.x
35	49.2	3.4	599	69	FR0019976	FR0019976 F.rubripe
36	49.2	3.4	997	69	CNS005ZE	CNS005ZE F.rubripe
37	48.2	3.4	656	64	AL119185	AL119185 DKFZp761N
38	48.2	3.4	509	40	FR0019949	FR0019949 F.rubripe
39	47	3.3	595	40	AA940565	AA940565 vz46e06.r
40	47	3.3	1100	70	CNS016KD	CNS016KD. Drosophil
41	46.6	3.3	243	31	AA298823	AA298823 Drosophil
42	46	3.2	371	22	R80054	R80054 y191f07.r1
43	46	3.2	436	38	AA773241	AA773241 af78g04.r
44	45.4	3.2	781	47	AU051007	AU051007 AU051007
45	45.2	3.2	1101	69	CNS006BP	CNS006BP Drosophil

RESULT 1

```
LOCUS       AI956440      527 bp      mRNA      EST      20-AUG-1999
DEFINITION  u177c02.y1 Sugano mouse kidney mk1a Mus musculus cDNA clone
            IMAGE:2136578 5' similar to gb:L14751 Mouse ATP receptor (MOUSE);
            mRNA sequence.
ACCESSION   AI956440
NID         g5749149
VERSION     AI956440.1  GI:5749149
KEYWORDS    EST.
SOURCE      Mus musculus
  ORGANISM  house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189116.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1001254
            Trace considered overall poor quality
            Seq primer: custom primer used
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1..527
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2136578"
                     /clone_lib="Sugano mouse kidney mk1a"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); site_2: DraIII (CACCANGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCGCTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a DraIII adaptor [TGTTGGCCRACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCANGTG). XhoI should
                     be used to isolate the cDNA insert. Size selection was
                     performed to exclude fragments <1.5kb. Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science). Custom primers for
                     sequencing: 5' end primer CTTCTGCCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGACA."
BASE COUNT      108 a     147 c    141 g    130 t    1 others
ORIGIN
```

ALIGNMENTS

```
Query Match            9.2%;   Score 130.8; DB 63;  Length 527;
Best Local Similarity  63.0%;  Pred. No. 2.3e-23;
Matches 218; Conservative 0; Mismatches 127; Indels 1; Gaps 1;

Qy    244 GAGGTGGGAGCTGGGACTGGTTGGTTGTTTGATGAGGATTTCAAGTTCATCCTGCTGCCTGTGGAGC 303
          |||| ||||  | || ||| | |||||| | ||| ||| ||   | |  ||   |  ||||| || |
Db     61 GAGCTGCCATACAAGTAGTAGATTAAGCAGGACTTCAAGTACATGTGCAGCCCANGTGC 120

Qy    304 TATGCAGTTGCTTTTGTTGGGCGTTGGGCGTTGGGCCTTAACGCCCCAACCCTAGTGGCTCTTCATC 363
          |||| | || |  || ||| |  ||  ||||  | || || || ||  ||  || |||  | |  ||
Db    121 TATTGGTTGAGGTGCGCGCGTGCTCGAGTTGGTTGCCTGAACGTCCGAGGGTCTTTATATGTTCCTA 180

Qy    364 TTCCGCCTCCGACCCTGGGGAGTGGCGAACGGGCCACCTACATGTTCCACCTGGCCATTGGTCAGAC 423
          ||| |   ||| | || || || || |||  ||  ||||  |||||  ||||| ||  ||| | |||
Db    181 TGCCGACTCATAAACCTGGAACGCCTACACCACACCTACAGTTTCACCTGGCGAGATTAGGAC 240

Qy    424 ACCTTGTATGTGCGTCGCGTGCGCCACCCTCATCTRACTATTATGCAGCCCACACACACACTGG 483
          ||| ||| | | ||| ||  || | |||  | |  ||  || || |||  ||  ||  ||||||
Db    241 TCTCTCTACCGCAGCCGTCCCCTGCCGTTCGCGTGTTATTACTTACGCCCGGATTGACCACTGG 300

Qy    484 CCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCGTTTTCTTTGGAACCTCTACTGC 543
          ||  | ||| | | ||| ||||| ||| |  || |||  ||| ||| | ||||||  |
Db    301 CCATTTAGCACGGGTGCTCGCAAGCTGGAGCGTAATCTCTTCTACACCAACCTCTACTGC 360

Qy    544 AGTGTCCTTTTCCTCACCTGCATCAGCGTGCACCGCTACCTGGGCA 589
          || |||| ||||| |||| ||  | |||| | ||||| ||||| ||||
Db    361 AGCATCCTCTTCCTCACGCTGAT-AGCGTGCACGGTGCCCTGAGTA 405
```

RESULT 2

```
LOCUS       FR0019937      586 bp      DNA      GSS      10-DEC-1997
FR0019937   F.rubripes GSS sequence, clone 037P03aG9, genomic survey sequence.
DEFINITION  F.rubripes GSS sequence, clone 037P03aG9, genomic survey sequence.
ACCESSION   AL012824
NID         g2679192
VERSION     AL012824.1  GI:2679192
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes
  ORGANISM  Fugu rubripes
            Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
```

```
source                    1..586
                          /organism="Fugu rubripes"
                          /db_xref="taxon:31033"
                          /clone="037P03"
                          /clone_lib="cosmid 037P03"

BASE COUNT        139 a      162 c      90 g      164 t      31 others
ORIGIN

Query Match          8.2%;   Score 117.2;  DB 69;  Length 586;
Best Local Similarity   64.6%;   Pred. No. 7.5e-20;
Matches  190;  Conservative   0;  Mismatches 103;  Indels   1;  Gaps   1;

Qy    253 CTGGACTGTTGGTTTGATGAGGATTCAAGTTCATCCTGCTGCTGTGTGAGCTATGCAGTT 312
          ||||||  ||| ||| ||| || ||||||| ||||||| || ||||||| || ||||||||
Db    180 CTCTACTGTAAATTCCAAGAGGACTTTAAATACAATTCTTCTCCCAGTCAGTTATGCCCTG 239

Qy    313 GTCTTTGTGTGCTGGGCTTGGGCCTTAACGCCCCAACCCTATGGCTCTTCAATCTTCCGCCTC 372
          ||||| || ||||| ||||| || |||| || ||| ||||| ||||||| ||  ||||||
Db    240 GTCTTTGTGTCTGGTTGGCCCTGAACGGGACGGGCCGCCCTTGTATGTGATGTCTTCCGCACA 299

Qy    373 CGACCCTGGGATCGAAGGGCCACCTACAGTTCCACCTGGCATTGTCAGACACCTTGTAT 432
          ||| || |||||| | || ||| ||||||| || ||||| ||||| |||||||| ||||
Db    300 AGGACCTGGAAGGCCATCCACTATCTACACATGTTCAACCTGACCATGTGTGACACCCTCTAC 359

Qy    433 GTGCGTGTCGCTGCCCACCCTCACTCACTATATTATGCAGCCCACAACCACTGGGCCCCTTGGC 492
          ||||| |||||| || ||| ||| || |||||||||| || |||| || ||||| || ||
Db    360 ATCCTCACACTACCCTTCCTCACTCATTACTACGGCGATGAAGAACGACTGGGCCC-TTCTC 418

Qy    493 ACTGAGAATCTGCAAGTTCGTCCCGCTTTCTTTTCTTTGGAACCTCTACTGCAGT 546
          || ||||||||| ||||| |||||| || || ||| ||| || |||| || ||||
Db    419 TGAACTCTCTGCAAGATCATTCGCTTTCTGTTTTTATGCCAACTTATACGGTATT 472

RESULT  3
LOCUS       AI344017      544 bp      mRNA        EST        08-APR-1999
DEFINITION  tc01e02.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2062586 3'
            similar to TR:000254 000254 PROTEASE-ACTIVATED RECEPTOR 3.
            ;contains TAR1.t3 MER22 repetitive element ;, mRNA sequence.
ACCESSION   AI344017
NID         g4081223
VERSION     AI344017.1    GI:4081223
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1798041.
```

```
FEATURES            Location/Qualifiers
     source         1..544
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="16"
                    /clone="IMAGE:2062586"
                    /clone_lib="NCI_CGAP_Co16"
                    /tissue_type="colon tumor, RER+"
                    /lab_host="DH10B"
                    /note="organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI;
                    Plasmid DNA from the normalized library NCI_CGAP_Co10 was
                    prepared, and ss circles were made in vitro. Following HAP
                    purification, this DNA was used as tracer in a subtractive
                    hybridization reaction. The driver was PCR-amplified cDNAs
                    from a pool of 5,000 clones made from the same library
                    (cloneIDs 1057416-1061255, and 1144584-1145351).
                    Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT         98 a      182 c      203 g      61 t
ORIGIN

Query Match          5.6%;   Score 80.2;  DB 45;  Length 544;
Best Local Similarity   52.2%;   Pred. No. 2.2e-10;
Matches  203;  Conservative   0;  Mismatches 183;  Indels   3;  Gaps   1;

Qy    285 CATCCTGCTGCTGGTGAGCTATGCAGAGTTGTTCTTTTGTGCTGGGCTTGGGCCTTAACGCCC 344
          |||||||| || ||  || |||  |||||  || ||  ||  || |||||| |||||||
Db    387 CAGGCTGGGTGCCCCGGCCCTCTTTATGGGGCTGGTCCTCGGCGGTCCTGGGGGTCCGCCAGGT 328

Qy    345 AACCCTATGGCTCTTCAATCTTCCGCCTCCGACCCTCCGGATGCAACGGGCCACCTACATGTT 404
          ||| |||||||| ||  |||| ||||| || |||| |||| |||  || ||| || ||||
Db    327 GGCGCGTGTGG---GTGCTGGCCACGGCAGCCACCCTGGGCGTCCCCTCCACCATGCTGCTGAT 271

Qy    405 CCACCTGGGCATTGTCAGACACCTTGTATGTGGTCGTGCGTGCCGGGGCCCCCGGGCCCGCCTACCA 211
          ||  ||||| || || ||||| ||  ||||||||| | || ||  ||||||| ||||||| |||
Db    270 GAACCTCGGACTGCGTGACCTCGTCGGCCCTGGCCCTGGCCTGCGCGGGATCGGCGCGCCTACCTACTATTA 464

Qy    465 TGCAGCCCACAACCACTGGCCCCTTTGGCACTGAGAATCTGCAAGTTCGTCCCGCTTTCTTTT 524
          || ||||||| ||  ||| || || | ||  ||||| |  || || || || ||| || ||||
Db    387 CCTGCGTGGCCAGCGGCTGGCCCTTCGGGGAGGCCGCCCTCCGGGCGCCGGCCGCGCACT 151

Qy    525 CTATTGGAAACCTCTACTGCAGTGTCCTTTTCCTCACCTGCCATCAGCGGTGCACCGCTACCT 584
          || ||||| ||||| || |||| || ||  || ||| || || || || || |||||||
Db    150 CTATGGTCACATGTATGGGTCCAGTGCTCGTCTGGCGGTCGCCTCAGCCTGGATCGGCTACCT 91

Qy    585 GGGGCATTCTSCCACCCCACTTCGGGGGCACTACGGCTGGGGCCCCCGCCCCTGCCAGGCCTTCT 644
          ||| | |||| | || || ||| |||| | |||| |||||||  ||| ||| || |||||||
Db    90 GGCCCCTGGGTGCCACCCGCGGGGCCCGGCGCCGGCCGCGGGGCCCTGGCGCGCCCTGGACT 31

Qy    645 CTGCCTGGCCAGTTTGGGTTGGTCGTAGCCG 673
          || ||| |||||  ||||||   || ||||
Db    30 CTGCATGGCTGCTTGGCCTCATGCGCGGCCG 2
```

```
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/iMAGE.html
```

```
FEATURES
    source          Seq primer: -400P from Gibco
                    High quality sequence stop: 356.
                    1..542
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:2461414"
                    /clone_lib="NCI_CGAP_Kid11"
                    /lab_host="DH10B"
                    /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                    a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                    Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                    prepared, and ss circles were made in vitro. Following HAP
                    purification, this DNA was used as tracer in a subtractive
                    hybridization reaction. The driver was PCR-amplified cDNAs
                    from a pool of 5,000 clones made from the same library
                    (cloneIDs 1322376-1323911, 1456007-1456775, and
                    1500552-1502855). Subtraction by Bento Soares and M.
                    Fatima Bonaldo."
```

```
BASE COUNT    175 a    123 c    135 g    109 t
ORIGIN
```

```
Query Match    5.3%;    Score 75.4;    DB 63;    Length 542;
Best Local Similarity 54.2%;    Pred. No. 3.7e-09;
Matches 174;    Conservative 0;    Mismatches 146;    Indels 1;    Gaps 1;

Qy    287  TCCTGCTGCCCTGTGAGCTATGCAGTTGTTCTTTGTGTGTCGGGCCTTAACGCCCCAA 346
           ||||| || || || ||||| |||||||||||||||| ||||||||| ||||  |
Db    517  TCTTCCTTCCAATTGTCTACACCATTGGTTGTTCGTTGGTGGGCTCGCCAAGTACGGCATGG 458

Qy    347  CC-CTATGGCTTCTTCATCTTCCGCCTCCGACCCTGGGATGCAACGGCCACCTACACGTTC 405
           ||  || |||||||| |||| |||||||||| |||||| || |||||| ||||||| |||
Db    457  CCTTTGTGGGTGCTTTTCTTTTCCGAATTAAGAGAAGCACCTTGCTGTGGATTTACATGGCC 398

Qy    406  CACCTGGCATTGTCAGACACCTTGTATGTCGTCTGCGCTGCCCACCCTCATCTACTATTAT 465
           ||||||||| ||||||| |||||||| ||||||| |||| || ||||||||| || |||||
Db    397  AATCTGGCATTGGCTGACCTCCTCTGTGTCATCTGGTTCCCCTTGAAGATTGCCTATCAC 338

Qy    466  GCAGCCCACAACCACTGGCCCTTTGGCCACTGAGATCTGCAAGTTCGTGCCCGCTTTCTTT 525
           |||| ||||||||||||||||| ||||||||||| || || | || ||||||||||||||
Db    337  ATACATGGCAACAACTGGATTTATGGGGAAGCTCTTTGTAATGTGCTTATTGGCTTTTTTC 278

Qy    526  TATTGGAACCTCTACTGCGTCCTTTTCCTCACCTGCATCAGCGGTTGCACGGCTACCTG 585
           ||||||||||||| ||||| |||| |||||||||||||| ||||||||| |||||||||
Db    277  TATGGCAACAATGTACTGTTCCAATCTCTTCATGAGACCTCGGCTGTGTGCCAGAGGTATTGG 218

Qy    586  GGCATCTGCCACCCACTTCGG 606
           |||||||||||||| |||||
Db    217  GTCATCGTGTGAACCCATGGGG 197
```

```
RESULT    5

LOCUS       R15256       466 bp    mRNA                    EST    13-APR-1995
DEFINITION  yf89o03.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:29866 5' similar to gb:M64749 G PROTEIN-COUPLED RECEPTOR RDC1
            HOMOLOG (HUMAN);, mRNA sequence.
ACCESSION   R15256
NID         g769529
VERSION     R15256.1  GI:769529
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Sep 21, 1992 this sequence version replaced gi:276439.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1857
            High quality sequence stops: 303 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: M13Rp1
            Insert Length: 1857   Std Error: 0.00
            High quality sequence stop: 303.
```

```
FEATURES
    source          1..466
                    /organism="Homo sapiens"
                    /sex="female"
                    /dev_stage="73 days post natal"
                    /clone_lib="Soares infant brain 1NIB"
                    /clone="IMAGE:29866"
                    /db_xref="taxon:9606"
                    /db_xref="GDB:402213"
                    /db_xref="GDB:40213"
                    /note="Organ: whole brain; Vector: Lafmid BA; Site 1: Not
                    I; Site 2: Hind III; 1st strand cDNA was primed with a Not
                    I - oligo(dT) primer [5'
                    AACTGGAAGAATTCGCGGCCGCCGGAATTTTTTTTTTTTTTTTT 3'];
                    double-stranded cDNA was ligated to Hind III adaptors
                    (Pharmacia), digested with Not I and directionally cloned
                    into the Not I and Hind III sites of the Lafmid BA vector.
                    Library went through one round of normalization. Library
                    constructed by Bento Soares and M.Fatima Bonaldo."
```

```
BASE COUNT    94 a    141 c    119 g    109 t    3 others
ORIGIN
```

```
Query Match    5.2%;    Score 74.2;    DB 21;    Length 466;
Best Local Similarity 59.1%;    Pred. No. 7e-09;
Matches 127;    Conservative 0;    Mismatches 88;    Indels 0;    Gaps 0;

Qy    383  ATGCAAACGGCCACCTACATGTTCCACCTGGCATTGTCAGACACCTTGTATGTGCTGCGC 442
           ||| ||| || |||||||||||| ||||||||||||||||||| || |||||| ||||
Db    49   ATGACACGCACTGCTACATCTTGAACCTGGCCATTGCCGACCTGGGGTTGTGCTCCACCA 108

Qy    443  TGCCCACCCTCATCTACTACTATTATGCAGCCCACAACCACTGGCCCTTTGGCACTGAGATCT 502
           || ||||||||||| || |  |||| |||||||||||||| ||||| |||||||| || |
Db    109  TCCCAGTCTGGGTGGTCAGTTCGTGCGAGCACAACCAGTGGCCCATGGGGAAGCTCACGT 168

Qy    503  GCAAGTTCGTCCGCCTTTCTTTCTATTGGAACCTCTACTGCAGTGTCCCTTTTCCTCACCT 562
           ||||||||| |||| |||||||||||||||||||||| ||||||  ||  |||||||||
Db    169  GCAAAGTCACACACCTCATCTTCTCCAATCAACCTCTTCGGCAGCATTTTCTTCCTCACGT 228

Qy    563  GCATCAGCGGTTGCACGGCTACCTGGGCGAATCTGCCAC 597
           ||| ||||| ||||||||||||| |  |||| |  |||
Db    229  GCATGAGCGGTGGACCGGCTACCTCCATCACCACCTAC 263
```

```
RESULT    6
```

```
LOCUS       AA296456         316 bp    mRNA            EST       18-APR-1997
DEFINITION  EST11038 Umbilical vein endothelial cells II Homo sapiens cDNA 5'
            end similar to vasoactive intestinal peptide receptor RDC1, mRNA
            sequence.
ACCESSION   AA296456
NID         g1948859
VERSION     AA296456.1  GI:1948859
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weldman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693170.
            Other_ESTs: EST11039 EST183921 THC105954
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 301869056
            Fax: 301869423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hg1/hg1.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..316
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):194234"
                     /db_xref="taxon:9606"
                     /clone_lib="Umbilical vein endothelial cells II"
                     /cell_type="endothelial cell"
                     /dev_stage="fetus"
                     /note="Organ: umbilical vein; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       67 a    106 c     72 g     71 t
ORIGIN
```

```
Qy   383 ATGCAAACGGCCACCTACACATGTTCCACCTGGCCATTGCAGAGACCTTGTATGTGCTCGC 442
         |||  ||| |||| ||||| || ||||||   |||||  ||  ||||||| ||| || |||
Db    48 ATGACAGCAGTCGCTACATCTTGAACCTGGCCATTCCCGACCTGTGGGTTGTCTCCTCACA 107

Qy   443 TGCCCACCCTTCATCTTACTATTATGCAGCCACCACAACACTGCCCTTTGGGACTGAGATCT 502
         ||| ||||| ||  ||||   |||| ||||| ||||  |||||||||||   ||| || ||
Db   108 TCCCAGTGCTGGGTGGTGCTCCGTCCGCAGCAACAACCAGTGGCCATGGGCCGAGCTCACGT 167
```

Query Match   5.2%;   Score 74.2;   DB 31;   Length 316;
Best Local Similarity   59.18;   Pred. No. 5.8e-09;
Matches   127;   Conservative   0;   Mismatches   88;   Indels   0;   Gaps   0;

```
RESULT  7
LOCUS       AA331167         261 bp    mRNA            EST       21-APR-1997
DEFINITION  EST35211 Embryo, 8 week I Homo sapiens cDNA 5' end similar to
            vasoactive intestinal peptide receptor RDC1, mRNA sequence.
ACCESSION   AA331167
NID         g1983628
VERSION     AA331167.1  GI:1983628
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weldman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407040.
            Other_ESTs: EST35210 EST179908 THC105954
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 301869056
            Fax: 301869423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hg1/hg1.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..261
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):133016"
                     /db_xref="taxon:9606"
                     /clone_lib="Embryo, 8 week I"
                     /dev_stage="embryo, 8 wks"
                     /note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       59 a     90 c     54 g     56 t      2 others
ORIGIN
```

```
Qy   503 GCAAGTTCGTCCGGCCTTTCTTGTTTCGTTATTGGAAACCTCTACTGCAGTGTCCGTTTCCTCACCT 562
         ||||  || || ||  ||| ||||| ||| ||||| ||||||||||||||  | ||||||||||||| |
Db   168 GCAAAGTCACACACCTCATCTTCTCATCAACACCTCTTCGGCAGACATTTTCTTCCTCACGT 227

Qy   563 GCATCAGGGTGGCACCGCTACCTGGGCACTGCCAC 597
         |||| |||||| |||||||| ||||||||||||||||  || ||
Db   228 GCATGAGCGTGGACGCCTAGCCTCTCACACCTAC 262
```

Query Match   5.2%;   Score 73.8;   DB 32;   Length 261;
Best Local Similarity   58.6%;   Pred. No. 6.7e-09;

Matches 126;   Conservative   0;   Mismatches   89;   Indels   0;   Gaps   0;

```
Qy   383  ATGCAACGGGCCACCTRCACTAGTTCCACCTGGCATTGTGACAACACCTTGTATGTGCTGTGCC 442
          ||||  ||||| || ||||  |||||| | |||||   |||||||||||||||||
Db   23   ATGACACGCACTGCTACATCTTGAACCTGCCATTGCCGACCTGTGGGTTGTGCCTCACCA 82

Qy   443  TGCCCACCCTCATCTACTACTATATGCAGCCCACACAACACTGGGCCCTTTGGCCACTGAGANCT 502
          |  ||  |||||||||| | || |||||||||||||||||  ||||||
Db   83   TCCCAGTCTGGGTGGGTCAGTCTGCGTGCGAGCACAACCAGTGGCCATGGGCAGCTCACGT 142

Qy   503  GCAAGTTCCGTCCGCTTCTTTCTTCTIATTGGAACCTCTACTGCGAGTGTCCTTTCCTCACCT 562
          ||||  |||||||  ||||| || |||||||||| ||||||||||| |||||
Db   143  GCAAAGTCACACACCCTCATCTTTCCATCAACCTCTTCGGCAGCATTTCTTCCTNACGT 202

Qy   563  GCATCAGCGTGCGACCCGCTACCTGGGGCAICTGCCAC 597
          ||||||| ||||| ||||||||||||| |||||
Db   203  GCATGAGCGTGGGACCGCTACCTNTCCATCAACTAC 237
```

RESULT   8

LOCUS   AA804531   599 bp   mRNA   EST   18-FEB-1998
DEFINITION ns28c04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184934 3',
            similar to TR:P79928 P79928 P2Y8 NUCLEOTIDE RECEPTOR. ;, mRNA
            sequence.
ACCESSION   AA804531
NID   g2873662
VERSION   AA804531.1   GI:2873662
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 599).
AUTHORS   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE   Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Jan 14, 1998 this sequence version replaced gi:1798299.
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            This clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
FEATURES   Location/Qualifiers
     source   1..599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="12"
                     /clone="IMAGE:1184934"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGGCGGCGCGCCGCTCATTTTTTTTTTTTTTTTT-
```

Query Match     5.0%;  Score 70.8;  DB 38;  Length 599;
Best Local Similarity    51.6%;  Pred. No. 5.9e-08;
Matches 159;  Conservative   0;  Mismatches 149;  Indels   0;  Gaps   0;

```
Qy   292  CTGCCTGTGAGGTATGGAGAGTGTGTCTTTGTGCTGGGGCCTTAACGACGCCACCTG 351
          |||| ||| ||| | |||||   |||||| |||||||||||||| ||||||||
Db   291  CTGCCCGTGGGGTGTTACTCCGCTGGGTGGGCGCGGTCAGCAATCCCGGGCAACCTCTCTCTG 350

Qy   352  TGGCTCTTCATCTTCCGCCTCCGACCCTGGGAGGCAACGGCCACCTACAGTTCCACCTG 411
          ||| |||||||||| |||||| |||  || ||||||| ||||| || |||||||| ||
Db   351  TGGGTGCTGTTGCCGGGCGCCCAGATCCCGTCGGGGTTCGGTGCATCCCAGGCGCAACCTG 410

Qy   412  GCATTGTCAGACACCTTGTATGTGGCTGTGCGCTGCGCCCACCCTCATCTACTATTATGCAGCC 471
          |||||||| |||| || ||||||| |||| ||||||| || |||||||||||||||| |||||
Db   411  ACGGTCACGGACCTGATCTGGGCCCAGGCGGTTGCCTTTCCAAAATCTACTACCATTGCAAC 470

Qy   472  CACAAACAACTGGCCCTTTGGCACTGAGANTCTGCAAGTTCGTCGCGCTTTCTTTTTCTATTGG 531
          | |||||||||||| || | || ||||||||||||||  ||| | | |||||||||||||
Db   471  CGCCACCACTGGGTATTCGGGGTGCTGCTTTGCAACGTGGGTGACCGTGGCCTTTTACCGA 530

Qy   532  AACCTCTACTGCGAGTGTCCTTTTCCTCACCTGCCATCAGCGTGCGACCGGCTACCTGGGGCATC 591
          |||||||||| | ||||||||| |||||||||||||||| ||| ||||||||||||||||||
Db   531  AACATGTAGTTCAGCATNCTCAGCATGAGNCTGTATCAGCGTGGAGGGCTTCCTTGGGGGGTC 590

Qy   592  TGCCCACCC 599
          || ||||||
Db   591  TGTAACGC 598
```

RESULT   9

LOCUS   AL048512   592 bp   mRNA   EST   29-SEP-1999
DEFINITION DKFZp586M0824_r1 586 (synonym: hutel) Homo sapiens cDNA clone
            DKFZp586M0824, mRNA sequence.
ACCESSION   AL048512
NID   g5935540
VERSION   AL048512.2   GI:5935540
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 592).
AUTHORS   Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
TITLE   EST (Duesterhoeft, et al.)
JOURNAL   Unpublished (1999)
COMMENT   On Apr 30, 1999 this sequence version replaced gi:4727652.
            Contact: Duesterhoeft A
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..592
                     /organism="Homo sapiens"
```

BASE COUNT   195 a   175 c   128 t   2 others
ORIGIN   99 a   175 g   128 t   2 others

3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

Insert Length: 1986   Std Error: 0.00
Seq Primer: -40m13 fwd. ET from Amersham
High Quality sequence stop: 420.

RESULT 10

LOCUS   AA177828   529 bp   mRNA   EST   16-FEB-1997
DEFINITION mt07f02.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:620379
            5' similar to TR:G663021 G663021 PROTEINASE ACTIVATED RECEPTOR 2.
            ;, mRNA sequence.
ACCESSION   AA177828
NID   g1759090
VERSION   AA177828.1 GI:1759090
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 529)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Thesing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1393684.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.

Query Match   53.7%; Score 70.8; DB 64; Length 592;
Best Local Similarity   5.0%; Pred. No. 5.9e-08;
Matches 189; Conservative   0; Mismatches 159; Indels   4; Gaps   2;

Qy   257 ACTGTTGGTTTGATGAGGATTTCAAGTTCATCCTGCTCGTGGACCTATGCCAGTTGTCTCT 316
           ||| | | | | ||| | ||| | || | || | || || || | || |
Db   229 ACTGCTTCTATAATGACTCCTTTAAGTACACTTGTATAGGGTGCAGTTCACCATGGTGT   288

Qy   317 TTGTGTCCTTGGGCTTGGGCCTTAACGCCCCAACCCTATGGCTCTTCATCTTCCGCTCCGAC 376
           |||||| || | || || |||| | | | ||||| | | ||
Db   289 TTGTGTCTTGGGTTAATATCCAAATTCGTTGTGCCATATACAATTTCATCTGCTCCTCAAAG 348

Qy   377 CCTGGGATGCAAACGGCCACCTACACATGTTCCACCT-GGCATGTGTCAGACACCTTGTATGTG 435
           | || | | |||| | |||| | | | | ||| ||||| | ||| | ||
Db   349 TCCGAAAATGAAAACTACAACTTACAICATTAACTTGGGCAATGTCAGAGACTTGCTTTGTT   408

Qy   436 CTGTCGGCTGCCCACCCTCAICTCACTACATAITATGCAGACCCACAACCACTGGCCCTTGGCACT 495
           | || ||| || | || || || | | || | || ||| |
Db   409 TTTACTTACCTTCAGGATTTTTACTTCACACACGGAAT---TGGCCATTGGAGAT   465

Qy   496 GAGACATGCAAGTTCGTCCGGCCTTTCTCTTATTGGAAACCTCCTACTGCAGTGTCCTTTC 555
           || || | |||| | | ||||||| || | | |||||| || | |||| || | ||
Db   466 TTACTTTGTAAGAATTTCTGTGATGCTGTTTTATACAACATGTACGGAAGCATTCTGTTC   525

Qy   556 CTCACCTGCATCAGCCGTGCACCGCTACCTGGGCATCTGCCACCACTTCGGG 607
           | || | | | || ||| |||| | | || |||| ||| |||| |
Db   526 TTAACCTGTRANTAGTGTTAGARTCTGGCAAATTGTCNACCCATTTAAGG 577

BASE COUNT   172 a   106 c   112 g   199 t   3 others
ORIGIN

FEATURES   Location/Qualifiers
source   1..529
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone_lib="Soares mouse 3NbMS"
                /clone="IMAGE:620379"
                /sex="male"
                /dev_stage="4 weeks"
                /tissue_type="Spleen"
                /lab_host="DH10B"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5'
                TGTTACCAATCTGAAGTAGGGGAGCGGGCGGGCGGTTTTTTTTTTTTTTTTT
                3']; double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. RNA
                provided by Dr. Bertrand Jordan. Library went through
                three rounds of normalization, and was constructed by
                Bento Soares and M.Fatima Bonaldo."

RESULT 11

LOCUS   AA611127   650 bp   mRNA   EST   01-OCT-1997
DEFINITION vo61f08.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:1054407 5' similar to gb:D13814 TYPE-1A ANGIOTENSIN II
            RECEPTOR (HUMAN);, mRNA sequence.
ACCESSION   AA611127
NID   g2461325
VERSION   AA611127.1 GI:2461325
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus

MGI:381203
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 494.

Query Match   50.3%; Score 70; DB 29; Length 529;
Best Local Similarity   4.9%; Pred. No. 8.9e-08;
Matches 200; Conservative   0; Mismatches 195; Indels   3; Gaps   1;

Qy   290 TGCTGCCTGTGAGCTATGCAGTTGTCTTTTGTGCTGGGCTTGGGCCTTAACGCCCCACCC 349
           || ||| |||| | | ||||||| |||| || |||| ||||| || ||| || ||
Db   22 TGATACCTGCCARCTACATCCTGCTGTTTGTGGTTGGGTGTGTTACCAGCGCAACACATGACC   81

Qy   350 TATGGCTCTTCATCTTCCGCTCCGACCCTCGGGGATGGCAAAGGGCCACCACTTCTTCCACC 409
           || |||||||| | | || |||| | | | ||| | ||| || ||| | || | ||
Db   82 TGTGGAAAACTCTCCTTAAGGACCAAAAATCCATCAGT---CTGGTGCATCTTTCACACCAACC   138

Qy   410 TGGCATGTCAGACACCTTGTATGTGCTGTCGGCTGCCCACCCTCATCTTACTAITTATGCAG 469
           ||||||||||| | ||| | || | ||| || || ||| | || || | |||
Db   139 TGGCCATGCGCAGATCTCCTTTCTGTGTGCACACTGCCATTTAAGATCGCCTACCACTCCA   198

Qy   470 CCCACAACCACTGGCCCTTGGCACTGAGACATGCAAGTTCGTCCGGCCTTTCTCTTATTCTATT 529
           || |||||| || ||| | | || || || ||||| | | |||||| | | |
Db   199 ATGGCAACAAACTGGGTARTTGCCGAGGTCACCTGCCCGGATCACCACGGTTGTTTTCTACG   258

Qy   530 GGAACCTCTACTGCAGTGTCCTTTCTCCTCACCTGCATCAGCCGTGCACCGCTACCTGGGCA 589
           |||| ||| | ||| | || | | | ||| | | | | ||| ||| || ||| |
Db   259 GCAACANGTACTGCGCGTATCCTGATCCTCACCTGCAATGGGCATCAACACGGCTACTGGCCA   318

Qy   590 TCTGCCACCACTTCGGGCACTACGCGTGGGGGCGCGCCCTTCGCCTTCGCCGAGGCCTTCTCGCC 649
           ||| | | || | ||| || || | || || || | ||| || | ||| | | |
Db   319 CGGCTCACCCTTTCACATACCAGAAGCGTGCCCAAACGGAGCTTCTCCATGCTCATTGTGTG 378

Qy   650 TGGCAGTTTGGTTGGTTCGGTAGCCGGCGTGCCCTCGTGCCC 687
           |||| | |||||||||| | || | || ||| || ||
Db   379 GCATGGTGTGGGTCANTGGTTTTCTTATACATGCCTGCCC 416

BASE COUNT   110 a   168 c   107 g   144 t
ORIGIN

REFERENCE   1 (bases 1 to 650)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1404829.

             Contact: Marra M/Mouse EST Project
             Washu-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:585983
             Seq primer: -28m13 rev2 ET from Amersham
             High quality sequence stop: 501.
FEATURES   Location/Qualifiers
     source   1..650
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1054407"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCGGCGGCGAATGGTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT   144 a   183 c   134 g   189 t
ORIGIN

Query Match   4.6%; Score 66; DB 36; Length 650;
Best Local Similarity   53.0%; Pred. No. 1e-06;
Matches 141; Conservative   0; Mismatches 125; Indels   0; Gaps   0;

Qy   407 ACCTGGCATTGTCAGAACACTTGTATGTGCTGTCCGCGTTGGT   466
          || ||| || ||||| |||| || ||| |||||||||||
Db   11 ATTCCCCCTGGATGGACTTATGCTTTTTGCCTGTTGCCTGATCTACTACCG   70

Qy   467 CAGCCCACAACCACTCGGCCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTTCTTTTTCT   526
          || | |||||| |||| | |||| ||| || |||| | ||||| ||| ||
Db   71 CTATGGAATACCGCGTGGCCCTTCGCCAATCACCTATGTAAGATCCTTCGCGCCACGGTCA   130

Qy   527 ATTGGAACCTCTACTGCAGTGTCCTTTTTCCTTCACCTCACTACGACGTGCACCCTACCTGG   586
          |||| ||| | | | ||| ||||||||| | |||||| || ||| ||| |||| ||
Db   131 GTTTCAAACCTCTACCGCCAGCCGTGTTCCGTGCTCAGCAGCGACGTCGATCGATCGACCTACCTGG   190

Qy   587 GCACTCGCCACCCACTTCGGGCCACTACGCTGGGGGCCGCCCCGCCCTCGCCGCCTTGCTTCTCT   646
          ||||||| || |||| |||| ||||| ||| |||||| ||| |||| || ||||| ||
Db   191 CCATTGTCCACCCGATGAAGTCTCGCCTCCCGCCGCACGATCGCTGGTGCCAAAGTCACCT   250

Qy   647 GCCTGGCAGTTGGTTGGTTCCTAGCC   672
          || |||||| ||||| || ||
Db   251 GCATCATCATCTGGCTGATGGCTGGC   276

───────────────────────────────────────

RESULT 12
AI607213
LOCUS   AI607213   476 bp   mRNA   EST   21-APR-1999
DEFINITION vo61f08.y1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:1054407 5' similar to gb:D13814 TYPE-1A ANGIOTENSIN II
            RECEPTOR (HUMAN).; mRNA sequence.
ACCESSION   AI607213
NID   g4616380
VERSION   AI607213.1 GI:4616380
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            house mouse.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 476)
AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE   The WashU-NCI Mouse EST Project 1999
JOURNAL   Unpublished (1999)
COMMENT   On Jun 5, 1998 this sequence version replaced gi:3189257.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:585983
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation).
            Seq primer: -40RP from Gibco
            High quality sequence stop: 465.
FEATURES   Location/Qualifiers
     source   1..476
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1054407"
                     /map="920H09; 1; 1p13.1-1q21.3"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCGGCGGCGAATGGTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT   96 a   146 c   97 g   137 t
ORIGIN

Query Match   4.6%; Score 66; DB 48; Length 476;
Best Local Similarity   51.7%; Pred. No. 9e-07;
Matches 150; Conservative   0; Mismatches 140; Indels   0; Gaps   0;

Qy   407 ACCTGGCATTGTCAGAACACTTGTATGTGCTGTCCGCGTTGGT...CCACCCCTCATCACTACTATTATG   466
          || ||| || ||||| |||| || ||| |||||| || | | |||| | |||| ||| |

RESULT 13

```
LOCUS       AA205847          816 bp    mRNA                     EST         12-MAR-1998
DEFINITION  zg50c02.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
            clone IMAGE:645026 5' similar to gb:M64749 G PROTEIN-COUPLED
            RECEPTOR RDC1 HOMOLOG (HUMAN);, mRNA sequence.
ACCESSION   AA205847
NID         g1801218
VERSION     AA205847.1  GI:1801218
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Thesing,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 24, 1995 this sequence version replaced gi:634356.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2201    Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1..816
                     /organism="Homo sapiens"
                     /db_xref="GDB:5215277"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:645026"
                     /clone_lib="Stratagene neuroepithelium (#937231)"
                     /dev_stage="Ntera-2/D1 neuroepithelial cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned undirectionally. Primer: Oligo dT; NT2
                     cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24
                     hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3'-3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      238 a    193 c    197 g    181 t    7 others
ORIGIN
```

```
Qy     11 ATCTGCCCTGGGCTGACTTATGCTTTTGCGACTTTGCTGTGTGGGGCTGTGTCTATACCG      70
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 CAGCCCAACACCACTGGGCCCTTGGCACTGAGGATCTGCAAGTTGTCCGCGCTTTCTTTTCT    526
```

```
Qy     71 CTATGGAATACCGGCTGGCCCTTCGGCAATCACCTATGTAAGATCGCTTCGGCCAGCGTCA     130
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 ATTGGAACCCTACACGAGTGTCCTTTTCCCACCGTGGCATCAGGCGTGCACCGCTACCTGG     586
```

```
Qy    131 GTTTCAACCTCTACGCCAGCCGTGTTCCCGTGCTCAGCATGCATCGATCCTACCTGG     190
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    587 GCATCTGCCACCCACTACGGCTGGGGCCCCGCCGTGGGCCGGCCTTCTCT     646
```

```
Qy    191 CCATTGTCCACCCGAGGAAGTCTCGCCCTCCCCGCACGATCCTGGGGCCAAAGTCACCT     250
          |||||||||||||||||||||||||||||||
Db    647 GCCTGGCAGTTTGGTTGGGTCGTTAGCCGGCGGCTGCCTGCCACCTTC     696
```

```
Qy    251 GCATCAATCATCTGGGCTGGATGGCTGGCGTTGGCCCAGTTTGGCCAGCCGTCATC     300
```

RESULT 14

```
LOCUS       AI739295/c        570 bp    mRNA                     EST         18-JUN-1999
DEFINITION  w130003.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2391725 3'
            similar to TR:O76067 O76067 PROTEASE-ACTIVATED RECEPTOR 4. ;, mRNA
            sequence.
ACCESSION   AI739295
NID         g5101276
VERSION     AI739295.1  GI:5101276
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189094.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1..570
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2391725"
                     /clone_lib="NCI_CGAP_Co16"
                     /tissue_type="colon tumor, RER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI;
```

```
Best Local Similarity  52.3%;
Matches 168; Conservative   0; Mismatches 152;  Indels   1;  Gaps   1;
                                                Pred. No. 1.3e-06;
```

```
Qy    278 TCAAGTTCATCCTCGTCCGCCTGTGGAGCTATGCAGGTCAGTTGTCTTGTCGTTGGGGCTTTA    337
          |  ||||                |||    ||      || |  ||| ||| |||||
Db    237 TCCTGCTCTTACACGGCTCCCTTCATTTACATTTTCATCTTCGTCATGCGGCATGATTGCCA    296
```

```
Qy    338 ACGGCCCCAAACCTATGGCTCTTCATCCTTCCGCCTCCGACCCCTGGGGGATGCAAACGGCCCACT    397
          || ||       ||||       |||          |||   ||  || ||
Db    297 ACTCCGTGGGTGGGTCTGGGTGAATATCCAAGGCCAAGACCACAGACGGCTATGACACGGCACTGCT    356
```

```
Qy    398 ACAATGTTCCACCTGGCATTGTGCAGACACCTTGTTATGTGCTGTGTGGCTGCCACCTCATCT    457
          |||||| || || || |||||| |  |||||| |    ||| ||||| ||  ||||||
Db    357 ACARCTTGAAACCTTGGCCATTGGCGACCTGTGTGGGGTTGTCGTCACCAATCCCAGTGTGGG    416
```

```
Qy    458 ACTATTATGCAGCCCACAACCACTGGCCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCCT    517
          ||||||| |||||||||  ||| ||||  |  || |||||||||    |||||  ||  ||
Db    417 TCAGTCTCGTGCAGACAACCAGTGGCCCATGGGAAGCTCACGTGCAAAGTGCACACACC    476
```

```
Qy    518 TTCTTTTCTATTGGAAACCTC-TACTGCAGTGTCCCTTTTCCTCACCTGCATCAGCGTGCAC    576
          || ||||                |||||||| ||                ||||| ||  ||
Db    477 TCATCTTCTCCATCAAACCTCTTTCGGCAGCATTTTCTTCCTCACGTGCATGCATGAGCGTGGAC    536
```

```
Qy    577 CGCTACCTGGGCATCTGCCAC    597
          ||  |||||       ||||
Db    537 CGTAACCTCTCCATCACCTAC    557
```

COMMENT   Plasmid DNA from the normalized library NCI_CGAP_Co10 was
            prepared, and ss circles were made in vitro. Following HAP
            purification, this DNA was used as tracer in a subtractive
            hybridization reaction. The driver was PCR-amplified cDNAs
            from a pool of 5,000 clones made from the same library
            (cloneids 1057416-1061255, and 1144584-1145351).
            Subtraction by Bento Soares and M. Fatima Bonaldo.
            "
BASE COUNT   101 a   192 c   211 g   65 t   1 others
ORIGIN

Query Match   4.68; Score 65.4; DB 51; Length 570;
Best Local Similarity 51.3%; Pred. No. 1.4e-06;
Matches 178; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

```
Qy  285 CATCCTGCTGCCTGTGAGCTATGCAGTTGTGTCGTGGTCGGCTTGGGCCTTAACGCCCC 344
            |||  ||| |||||||||  ||||||| |||| |||||  ||| ||||||||  |||
Db  344 CAGGCTGGTGCCCCGCCCTCTANGGGCCTGGCTGGTTGGGGTGGGGGCGGCGCCAATGGGCt 285

Qy  345 AACCCTATGGCTCTTCATCTTCCGCCCTCCGACCCTGGCAAACGGCCGCCACCTACANGTT 404
            | ||| ||||||||||||||   |  ||  ||||||    ||||||||| ||||| |||
Db  284 GGCGCTGTGG---GTGCTGGCCGCAGGCACCTCGGCGTGCCCTGCCCTCCACCATCTGCTGAT 228

Qy  405 CCACCTGGCATTGTCAGACACCTTGTATGTGCGTCGGTCGCTGCCCACCCTCATCTACTATTA 464
            |||||||  |   | || | ||| |||| | || ||| |||||||| ||||||||||||
Db  227 GAACCTCGCGACTGCTGCACCCTCTGCTGCCCTCGGGGCCTGGGGGATCGCACCTACCA 168

Qy  465 TGCAGCCCACAACCACTGGGCCCCTTTGGCACTGAGATCTGCAAGTTTGCTCCCGTTTCTTTT 524
            |||||||  | || ||||| ||| ||||||| |||||| ||||| || ||| || | |||
Db  167 CCTGCGGTGGCCCAGGCGCCTTCGGGGAGGGCCCGCCCTGCCACCGCCCGCCGCCGCCACT 108

Qy  525 CTATTGGAACCTCTACTGCAGTGTCCTTTTCCTCACCTGCATCACGGGTGCCACCGCCTACCT 584
            ||||  |  | ||| || |||| |  ||| | |  ||||||| || ||  || || ||||||
Db  107 CTATGGTCACANGTANGGCTCAGTGTGCTGCTCGTCGGCGGGGATCGGCTACCT 48

Qy  585 GGGCATCTGCGCCACCACTTCGGGCCACTACGGCTCGGGGGGCGGCGCCCTCGCC 631
            |||| ||| |||||||   |  ||  ||| | || | || ||| ||||| |||||
Db   47 GGCCCTGGTGCACCCCCTGCGGGCCGCGGCCGCCCTGCGCGTGGGCCGGCGCC 1
```

RESULT 15
N93987

LOCUS   N93987   391 bp   mRNA   EST   05-APR-1996
DEFINITION za66d09.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:297521 5' similar to SW:GCR1_HUMAN P43657 RB INTRON ENCODED
            PUTATIVE G-PROTEIN COUPLED RECEPTOR. ; mRNA sequence.
ACCESSION   N93987
NID   g1266296
VERSION   N93987.1 GI:1266296
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 391)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., and
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P., and
            Wilson,R.
TITLE   The Washu-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   On Oct 18, 1995 this sequence version replaced gi:1025573.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET

High quality sequence stop: 282.
FEATURES   Location/Qualifiers
     source   1..391
                     /organism="Homo sapiens"
                     /db_xref="GDB:1242443"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:297521"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCGAAGTGGGACGGCCGCGACCGGCCAATTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library. Soares
                     fetal heart NbHH19W."
BASE COUNT   104 a   77 c   79 g   128 t   3 others
ORIGIN

Query Match   4.48; Score 62.6; DB 25; Length 391;
Best Local Similarity 52.3%; Pred. No. 6.1e-06;
Matches 162; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

```
Qy  378 CTGGGANGCAACGGCCACCTACACATGTTCCACCTGGCATTGTCAGACACACTTGTATGTGCT 437
            |  |  |  ||  | || | |||  |  || |||||  | || | |||| ||||| | |
Db    4 CCGAAATGAAAACTACAACTTACNTGATTAACTTGGCAANTGTCAGAGACTTGCTTTTTGTTT 63

Qy  438 GTCGGCTGCCCACCCTCATCACTATATTATGCAGCCCACAACACTGGGCCCTTTGGGCACTGA 497
            | |   |  || ||  | |  || ||||  | |   ||||  ||||| ||| || ||| |
Db   64 TACTTTACCCTTCAGGAATTTTTTTACTTCACAACACGGA---ATTGGCCATTTGGAGAATTT 120

Qy  498 GATCTGCAAGTTCGTCCGCCTTTTCTTTTCTTCTTATTGGAACCTCTACTGCAGTGTGTCCTTTTCCT 557
            ||||| || ||  ||    | || || ||| | ||| | ||  |  ||| |  ||||| |  |||
Db  121 ACTTTGTTAAAGATTTCGTTGGATGCTGTTTATACCACACATGTACGGAAGCAATTCTGTTCTT 180

Qy  558 CACCTGCATCAGCGTGCACGGTACCTGGGGCATCTGCCACCCACTTCGGGGCACTACGGCTG 617
            ||| || ||||   |  | || |  |  ||||  |||| ||| ||| || | ||| || |
Db  181 AACCTGTATTAGTGTAGAATCGANTTCTGGCAANTTGTCTACCCATTTAAGTCAAAAGACTCT 240

Qy  618 GGGCCCGCCCTGCCCTGCCAGGCCCTTCTTCTGCCTGGGCAGTTTGGTTGGTGCGCGCCGCTG 677
            ||| ||||| |||| |||||   || | | ||| |||||||| |||| ||| |  | |||
Db  241 AAGAACACAAAAGAAANTGCAAAGAATTGTTGCACTGGCGTGGGTTAACTGTGATCGGCGGAG 300

Qy  678 CCTCCGTGCCC 687
            || ||||||
Db  301 GAAAGTGCAC 310
```

Search completed: December 29, 1999, 04:26:25
Job time: 2546 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 04:51:36 ; Search time 104.56 Seconds
(without alignments)
3419.324 Million cell updates/sec

Title:
Perfect score: US-09-077-173A-1
Sequence: 1429
   1 AAGGGAGCTTGGGTAGGGGGC..............GCCACGGAGGCAGGGGATATC 1429

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database :   N_Geneseq_36:*

Word size :   0

Number of hits that pass the threshold :   623170

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1424.2	99.7	1429	1	T74321	Human P2Y4 recepto
2	350.4	24.5	1842	1	Q88134	Human P20 receptor
3	218.6	15.3	984	1	T18368	Human placenta G-p
4	210.8	14.8	984	1	T18367	Mouse pancreas G-p
5	210.6	14.7	984	1	T4730	Human placenta pur
6	210.6	14.7	984	1	T70000	Human placenta pur
7	128.4	9.0	1901	1	Q6178	P2U purinergic rec
8	128.4	9.0	2453	1	T44092	Seven transmembran
9	128.4	9.0	1901	1	V18356	Human G-protein th
10	107.6	9.5	1901	1	Q37107	Human R12 seven tr
11	107.6	7.5	1679	1	Q80522	New platelet facto
12	107.6	7.5	1679	1	Q99009	Human lymphocyte P
13	104	7.3	2427	1	X07369	Chemokine superfam
14	98.6	6.9	1477	1	Q84557	Murine P2Y11 recept
15	98.6	6.9	2732	1	Q84559	Murine C140 recept
16	98.6	6.9	1477	1	T32036	Murine C140 recept
17	98.6	6.9	2732	1	T32038	Murine C140 recept
18	96.4	6.7	1670	1	V26557	Human IP-10/Mig re
19	94.8	6.6	1876	1	T72800	Human G-protein ch
20	94.6	6.6	2216	1	Q66556	Murine delta opioi
21	93.8	6.6	1435	1	X06948	HEOAD54 cDNA parti
22	93	6.5	1821	1	Q65700	Sequence of murine
23	93	6.5	2272	1	Q75927	Sequence of murine
24	93	6.5	2218	1	V49253	Mouse delta opiate
25	92.2	6.5	1597	1	V07250	DNA encoding a G-p
26	91.2	6.4	1255	1	Q84558	Human C140 recepto
27	91.2	6.4	1414	1	Q84560	Human C140 recepto
28	91.2	6.4	1255	1	T32037	Human C140 recepto
29	91.2	6.4	1414	1	T32039	Human C140 recepto
30	91.2	6.4	1594	1	X06947	cDNA clone HEOAD54
31	90.8	6.4	1872	1	Q6176	Seven transmembran
32	90.8	6.4	1872	1	V18354	Human R20 seven tr
33	90	6.2	1301	1	T33905	Human G-protein co
34	85.2	6.0	1296	1	Q45657	Human somatostatin
35	84.2	5.9	1312	1	T93367	Mouse thrombin rec
36	83.4	5.8	1361	1	X16671	G-protein coupled
37	81.4	5.7	2864	1	V4374	Human G-protein co
38	81.4	5.7	1395	1	V4400	Human G-protein co
39	80.6	5.6	1110	1	Q29155	Pituitary somatost
40	80.2	5.6	1080	1	T30385	Probe based on ang
41	79	5.5	1200	1	Q30011	Sequence encoding
42	79	5.5	1200	1	Q9949	Recombinant high a
43	79	5.5	1029	1	T68863	Pig growth hormone
44	79	5.5	1029	1	T69755	Swine growth hormo
45	78.8	5.5	3480	1	Q28568	Human thrombin rec

ALIGNMENTS

RESULT   1
T74321

ID   T74321 standard; DNA; 1429 BP.
AC   T74321;
DT   31-MAR-1998   (first entry)
DE   Human P2Y4 receptor genomic DNA.
KW   Receptor; P2Y4; pyrimidine binding; uridine triphosphate; UTP; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   181..1278
FT   /*tag= a
FT   /product= P2Y4 receptor
PN   WO9719170-A1.
PD   29-MAY-1997.
PF   21-NOV-1996; BE0123.
PR   21-NOV-1995; EP-870124.
PA   (EURO-) EUROSCREEN SA.
PI   Boeynaems J, Communi D, Parmentier M, Pirotton S;
DR   WPI; 97-402177/37.
DR   P-PSDB; W23606.
PT   Receptor having preference for pyrimidine over purine nucleotide(s)
PT   - especially uridine tri:phosphate, agonist and antagonists of which
PT   are useful in treatment of cystic fibrosis
PS   Claim 14; Figure 1; 56pp; English.
CC   This sequence encodes a novel human P2 receptor, P2Y4, which has a
CC   preference for pyrimidine binding, especially uridine triphosphate.
CC   This receptor could be used to screen for novel drugs which
CC   specifically bind to it. Transgenic animals could be used to determine
CC   the physiological effects of expressing varying levels of the receptor
CC   or to identify novel agonists or antagonists. The agonists and
CC   antagonists of human P2Y4 may be used, e.g., in treatment of cystic
CC   fibrosis.
SQ   Sequence   1429 BP;   242 A;   451 C;   379 G;   357 T;

Query Match   99.7%;   Score 1424.2;   DB 1;   Length 1429;
Best Local Similarity   99.8%;
Matches 1426;   Conservative   0;   Mismatches   3;   Indels   0;   Gaps   0;

```
Qy      1 AAGGGAGCTTGGGTAGGGGCAGGGCTAGCCTGAGTGCAACCAGAGATGGCCTTCTGTCAGCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGGGAGCTTGGGTAGGGGCAGGGCTAGCCTGAGTGCACCAGAGATGGCCTTCTGTCAGCT 60

Qy     61 CTCCCTAGTGCTTCAACCACTGCTCTCCCTGCTCTACTTTTTGCTCCAGCTTCAGGGAT 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTCCCCAGTGCTTCAACCACTGCTCTCCCTGCTCTACTTTTTGCTCCAGCTTCAGGGAT 120

Qy    121 GGGGGGTGGGGCCAGGGAAATCCTGCCACCCTCACTTCTCCCCCTTCCCATCTCCAGGGGCC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGGGTGGGGCCAGGGAAATCCTGCCACCCTCACTTCTCCCCCTTCCCATCTCCAGGGGGCC 180

Qy    181 ATGGCCAGTACAGAGAGTCCTCCCTGTTGAGAGATCCCTAGGCCTTGGCCAGGGTCCTGGCAGC 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGGCCAGTACAGAGAGTCCTCCCTGTTGAGAGATCCCTAGGCCTCAGCCCCAGGGTCCTGGCAGC 240

Qy    241 AGTGAGGTGGGAGACCTGGTTGGTTGTTTTGGATGAGGATTTCATGCGTGCTCGGTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTGAGGTGGGAGACCTGGTTGGTTTTGGATGAGGATTTCATGCGTGCTCGGTG 300

Qy    301 AGCTATGCAGTTGTCTTGTCTTTGTGCTGGGGGCTTGGGGCCTTAACGCCCCACCCTATGGCTCTTC 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCTATGCAGTTGTCTTTGTTGTTGTGGGGGCTTGGGGCCTTAACGCCCCAACCCTATGGCTCTTC 360
```

```
RESULT 2
Q88134
ID    Q88134   standard; cDNA; 1842 BP.
AC    Q88134;
DT    29-NOV-1995   (first entry)
DE    Human P2U receptor gene.
DT    04-OCT-1994;  U11260.
PF    15-OCT-1993;  US-138137.
KW    Epithelial mucosa; mucus; cystic fibrosis; asthma; ss;
KW    chronic bronchitis.
OS    Homo sapiens.
FH    Key            Location/Qualifiers
cds                  57..1181
FT                   /*tag- a
PN    WO9510538-A.
PD    20-APR-1995.
PF    04-OCT-1994;  U11260.
PR    15-OCT-1993;  US-138137.
PA    (UMOR ) UNIV MISSOURI.
PA    (UYNC-) UNIV NORTH CAROLINA.
PI    Boucher RC,   Erb LJ,  Harden TK,  Lustig KD,  Parr CE;
PI    Sullivan DM,  Turner JT,  Weisman GA;
DR    WPI; 95-169967/22.
DR    P-PSDB; R72457.
PT    DNA encoding human P2U receptor and null cells expressing the
PT    receptors - for stimulating or inhibiting growth of cultures of
PT    mammalian cells, and for treating diseases of airway epithelial
PT    mucosa, e.g. asthma.
PS    Claim 3; Page 34; 47pp; English.
CC    The sequence is that of the human P2U receptor gene. The gene may
CC    be used to treat diseases of epithelial mucosal surfaces by
CC    enhancing or inhibiting mucus prodn. The gene may be used to treat
CC    cystic fibrosis, asthma and chronic bronchitis.
CC    See also Q88135-40.
SQ    Sequence 1842 BP;   364 A;   557 C;   528 G;   393 T;

Query Match       24.5%;   Score 350.4;   DB 1;   Length 1842;
Best Local Similarity   64.9%;                     Pred. No. 2.9e-87;
Matches  570;   Conservative   0;   Mismatches  296;   Indels  12;   Gaps  3;
```

```
Qy   243  TGAGGTGGAGCTGGACTGTTGGGTTTGATGAGGATTTCAAGTTCATCCTGCCTGTGGAG  302
          ||||  |||                 | | |  ||||| ||||| || | ||||||||||
Db   113  TGAGCTGGGGCTACAGGTGCCGCCGCTTCAACGAGGACTTGAAGTACGTGCCTGCCTGTGTC  172

Qy   303  CTATGCAGTTGTCTTTGTTGTGCTGGGCCTTGGGCCTTAAACGCCCCAACCCTATAGGCTCTTCAT  362
          ||| ||||        | |||||  || ||||  ||  |                |  || ||||||
Db   173  CTACGGCGTGGGTGTGCGGTCCTTGGGCGTGTCTGAAACGGCGTGGGCCCTTACATCTTCTT  232

Qy   363  CTTCCGGCCTCCGACCCCTGGGATGGCAAACGGCCACCTACACTACATGTTCCACCTGGCCATTGTCAGA  422
          ||||            |  ||||   || |||        ||||| |||||  || | ||||| || |
Db   233  GTGCCGCCCTCAAGACCTGGAATGCGTCACCACACATATAGTGTTCCACCTGGCTGTGTCTGA  292

Qy   423  CACCTTGTATGTGCTGTCGCGTGCCCACCCTCATCTACTATTATGCAGCCCACCACCACTG  482
          | ||  |     | || ||| | ||        | | |  |  | ||       || |||||
Db   293  TGCACTGTATGCGGGCGCCTCCCTGCCGTGGTCTATTACTACGCCGGCGGACCACTG  352

Qy   483  GCCCTTGGCACTGAGATCTGCAAGTTCGTCCCGCTTCTTTTTCTATTGGAACCTCTACTG  542
          ||||||       |  | |||  |||  |||  ||||         || ||   |||||
Db   353  GCCTTCAGCACGTGCTCTGCAAGCTCGTGGGCCGCTTCTTCTTCTACACCAACCTTTACTG  412

Qy   543  CAGTGTCCTTTTCCTCACCGTGCCATCAGCGTGCACCGGCCATCTGCCACCCACT  602
          |||  |||||       ||| ||||         |||   || | || |||||||
Db   413  CAGCATCCGTCTTCCTCACCGTCCGCATCAGCGTGCACCGGCGTCTCGGGCGTCTTTACGGCACCCACT  472

Qy   603  TCGGGCACTACGGTGGGCGCCCCTCGCCTCGCGAGGCCGTTCTTCTGCCTGGCAGTTTGTT  662
          |        |||||  ||||      ||   | || || |||  || |||| ||||| ||||
Db   473  GCGCTCCCTCGCCTGCGGGCGGCCGCCGGTACCGCTGCCGCCGGGCGTGTGGGT  532

Qy   663  GGTGGTAGCCGGCGCTGGCCCTCGTGCCCAACCTGTTCTTTGTCACAACCAGCAACAAAGGGGAC  722
          | |  |   |||| | || ||   ||  || ||| || || | || |  ||| ||| ||||
Db   533  GTTGGTGCTGGCCTTGCCAGGCCCGTGCTTCTACTTTGTCACCACCAGCGCGCGGGG--  591

Qy   723  CACCGGTCCTGTGCCATGACACCACTCGGCCGTGGAAGAGTTTGACCACTATGTGCACTTCAG  782
```

```
Qy   361  ATCTTCCGGCCTCCGACCCCTGGGATCGAACGGCACCTACATGTTCCACCTGGCATTGTCA  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  ATCTTCCGGCCTCCGACCCCTGGGATCGAACGGCACCTACATGTTCCACCTGGCATTGTCA  420

Qy   421  GACACCTTGTATGTGCTGTCGCGTGCCCACCCTCATCTACTATTATGCAGCCCACCACCAC  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  GACACCTTGTATGTGCTGTCGCGTGCCCACCCTCATCTACTATTATGCAGCCCACCACCAC  480

Qy   481  TGGCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTCCTGTTTCTATTGGGAACCTCTAC  540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  TGGCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTCCTGTTTCTATTGGGAACCTCTAC  540

Qy   541  TGCAGTGTCCTTTTCCTCACCGTGCCATCAGCGTGCACCGGCCATCTGCCACCCA  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TGCAGTGTCCTTTTCCTCACCGTGCCATCAGCGTGCACCGGCCATCTGCCACCCA  600

Qy   601  CTTCGGGCACTACGGTGGGCGCCCCTCGCCTCGCGAGGCCGTTCTTCTGCCTGGCAGTTTGG  660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  CTTCGGGCACTACGGTGGGCGCCCCTCGCCTCGCGAGGCCGTTCTTCTGCCTGGCAGTTTGG  660

Qy   661  TTGGTGGTAGCCGGCGCTGGCCCTCGTGCCCCAACCTGTTCTTTGTCACAACCAGCAACAAAGGG  720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  TTGGTGGTAGCCGGCGCTGGCCCTCGTGCCCAACCTGTTCTTTGTCACAACCAGCAACAAAGGG  720

Qy   721  ACCACCGGTCCTGTGCCATGACACCACTCGGCCGTGGAAGAGTTTGACCACTATGTGCACTTC  780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  ACCACCGGTCCTGTGCCATGACACCACTCGGCCGTGGAAGAGTTTGACCACTATGTGCACTTC  780

Qy   781  AGCTCGCCGGTCATGGGGCTGCTCTTTGGCGTGCCCGGCTGGACTCTTGTTGTTGCTAT  840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  AGCTCGGCCGGTCATGGGGCTGCTCTTTGGCGTGCCCGGCTGGACTCTTGTTGTTGCTAT  840

Qy   841  GGACTCATGGCTCGTCGTCGCCGTATCAGCAGCCCTTGCCACAGTCGTCGTTCTTCTCGC  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  GGACTCATGGCTCGTCGTCGCCGTATCAGCAGCCCTTGCCACAGTCGTCGTTCTTCTCGC  900

Qy   901  CTCCGCTCCTCTCCCACCATAGCTGTGGTGCTGCTGTCTTCGCTGCTTCGTGCCT  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  CTCCGGCTCCTCTCCCACCATAGCTGTGGTGCTGCTGTCTTCGCTGCTTCGTGCCT  960

Qy   961  TTCCACCATCACCCGCACCACCATTTACTACCTGGCCCAGGCTCGTTGGAAGCTGTCGCCGTA  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  TTCCACCATCACCCGCACCACCATTTACTACCTGGCCCAGGCTCGTTGGAAGCTGTCGCCGTA  1020

Qy  1021  CTGAACACATTGTCAACGTGGTCTATAPAAAGTGACTCGGCCCCTGGCCAGTGCCCAACAGCTGC  1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  CTGAACACATTGTCAACGTGGTCTATAPAAAGTGACTCGGCCCCTGGCCAGTGCCCAACAGCTGC  1080

Qy  1081  CTGGATCCTGTTGCTCTTACTTGCTCACTGGGACAAAATAATCGAGGTCAGGTCCGTCAGGTC  1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  CTGGATCCTGTTGCTCTTACTTGCTCACTGGGACAAAATAATCGAGGTCAGGTCCGTCAGGTC  1140

Qy  1141  TGTGGTGGTGGGGCAAGCCCCAGACCCCAGACGGCTGCTCCTCCTGGCCACTAGTGTCCCG  1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  TGTGGTGGTGGGGCAAGCCCCAGACCCCAGACGGCTGCTCCTCCTGGCCACTAGTGTCCCG  1200

Qy  1201  CCTGAGGATAGAGCTGCAGGTGGGCGGCCGCCACCCCCAGGACAGTAGTGCTGCTCTACTCCT  1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  CCTGAGGATAGAGCTGCAGGTGGGCGGCCGCCACCCCCAGGACAGTAGTGCTGCTCTACTCCT  1260

Qy  1261  AGGGCAGAGATAGATTGTTAACACGGGAAGCCGGGGAATGTGAGAGAAAAAGGGGGATGAGTGCAGG  1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  AGGGCAGAGATAGATTGTTAACACGGGAAGCCGGGGAATGTGAGAGAAAAAGGGGGATGAGTGCAGG  1320

Qy  1321  GCAGAGGTGAGGGAACCCAATAGTGATAACCTGGTAAGGTGCTTGCTTAGCAGTTTTGGT  1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  1321  GCAGAGGTGAGGGAACCCAATAGTGATAACCTGGTAAGGTGCTTCTTCCTCTTTTTCCAGGC  1380

Qy  1381  TCTGGAGAGAAGAGCCCTCACCCTGAGGGTTGCCACGGAGGAGGGATATC  1429
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  TCTGGAGAGAAGAGCCCTCACCCTGAGGGTTGCCACGGAGGAGGGATATC  1429
```

RESULT 3

ID   T18368 standard; cDNA; 984 BP.
AC   T18368;
DT   26-AUG-1996 (first entry)
DE   Human placenta G-protein coupled receptor protein cDNA phAH2-17.
KW   G-protein coupled receptor protein; G-PCRP; agonist; antagonist;
KW   cystic fibrosis; incontinence; diabetes; diagnosis; gene therapy;
ds.
OS   Homo sapiens .
PN   WO9605302-A1.
PD   22-FEB-1996.
PF   10-AUG-1995; J01599.
PR   11-AUG-1994; JP-189272.
PR   11-AUG-1994; JP-189273.
PR   11-AUG-1994; JP-189274.
PR   30-SEP-1994; JP-236356.
PR   30-SEP-1994; JP-236357.
PR   02-NOV-1994; JP-270017.
PR   28-DEC-1994; JP-326611.
PR   20-JAN-1995; JP-001177.
PR   16-MAR-1995; JP-057186.
PR   19-APR-1995; JP-093989.
PA   (TAKE) TAKEDA CHEM IND LTD.
PI   Fujii R, Fukusumi S, Hinuma S, Hosoya M, Ohgi K;
PI   Ohtaki T;
DR   WPI; 96-139698/14.
DR   P-PSDB; R91225.
PT   G-protein coupled receptor protein DNA and protein - also methods
PT   for isolating (antagonists for treatment of cystic fibrosis,
PT   incontinence and diabetes
PS   Claim 12; Page 269; 360pp; English.
CC   A cDNA clone (T18368), designated phAH2-17, codes for a novel human
CC   G-protein coupled receptor protein (G-PCRP) (R91225). It was obtd.
CC   by PCR amplification of a human placenta cDNA library using
CC   primers (see also T13910 and T13913) based on a mouse cDNA clone.
CC   Such cDNA sequences (see also T18366-67, Q18369-73 and T13901-09).
CC   may be used in the prodn. of recombinant G-PCRPs, in the prepn. of
CC   transgenic animals, and also for gene therapy.
SQ   Sequence 984 BP; 156 A; 361 C; 261 G; 206 T;

Query Match   494;
Best Local Similarity   55.2%;
Matches 494;   Conservative   0;   Mismatches 389;   Indels 12;   Gaps 3;

Score 218.6; DB 1; Length 984;
15.3%;   Pred. No. 5.2e-51;

RESULT 4

ID   T18367 standard; cDNA; 984 BP.
AC   T18367;
DT   26-AUG-1996 (first entry)
DE   Mouse pancreas G-protein coupled receptor protein cDNA pMAH2-17.
KW   G-protein coupled receptor protein; G-PCRP; agonist; antagonist;
KW   cystic fibrosis; incontinence; diabetes; diagnosis; gene therapy;
ds.
OS   Mus sp.
PN   WO9605302-A1.

```
PD  22-FEB-1996.
PF  10-AUG-1995; J01599.
PR  11-AUG-1994; JP-189272.
PR  11-AUG-1994; JP-189273.
PR  11-AUG-1994; JP-189274.
PR  30-SEP-1994; JP-236356.
PR  30-SEP-1994; JP-236357.
PR  02-NOV-1994; JP-270017.
PR  28-DEC-1994; JP-326611.
PR  20-JAN-1995; JP-007177.
PR  16-MAR-1995; JP-057186.
PR  19-APR-1995; JP-093989.
PA  (TAKE) TAKEDA CHEM IND LTD.
PI  Fujii R, Fukusumi S, Hinuma S, Hosoya M, Ohgi K;
PI  Ohtaki T;
DR  WPI; 96-139698/14.
DR  P-PSDB; R91224.
PT  G-protein coupled receptor protein DNA and protein - also methods
PT  for isolating (antagonists for treatment of cystic fibrosis,
PT  incontinence and diabetes
PS  Claim 12; Page 25-57; 360pp; English.
CC  A cDNA clone (T18367), designated pMAH2-17, codes for a mouse
CC  pancreatic beta-cell strain MIN6-derived G-protein coupled receptor
CC  protein (G-PCRP) (R91224). It was obtd. by PCR amplification of
CC  MIN6 cDNA using primers (T18347-48) based on the transmembrane
CC  domains of known G-PCRPs. Such cDNA sequences (see also T18366-73,
CC  and T13901-09) may be used in the prodn. of recombinant G-PCRPs,
CC  in the prepn. of transgenic animals, and also for gene therapy.
SQ  Sequence 984 BP; 163 A; 337 C; 255 G; 229 T;

Query Match      14.8%; Score 210.8; DB 1; Length 984;
Best Local Similarity  54.9%; Pred. No. 7.4e-49;
Matches 485; Conservative  0; Mismatches 387; Indels 12; Gaps  3;

Qy   270 TGAGGATTTCAAGTTCATCCTGCTGCCTGTGAGCTATGCAGTTGTCTTTGTCTGGGGCTTT 329
         |||||||||||| ||  ||   |  |||||||||  || ||  |||||||| |||   ||
Db    63 TGAGGATTTCAAGCGACTCGTTCAACCCCGGTATACTCGGTGTTGCTGGTGGCGTCGGGCT 122

Qy   330 GGGCCTTAACGCCCCAACCCTATGGCTCTTCATCTTCCGCCTCCGACCCTGGGATGCAAC 389
         |||||||||||| ||   |||| |||||||| | ||||||||||||||||||| ||||||
Db   123 GCCACTGAACANCTGCGTCAATCTGCGCATCCGCCCTCCGACCCTGACCCGGTTC 182

Qy   390 GGCCACCTACATGTTCCACCTGCGCATTGTCAGACACCTTGTATGTGCTCGCTGCCCAC 449
         |||  |||||||||       |||| | || |  ||||| || |||| || | |||||
Db   183 CGGTCGTGTACACCCTGAACCTGCCACTGCGGACCTGATGTATCCCTGTTCACTACCCCT 242

Qy   450 CCTCATCTACTATTATGCAGCCCACAACCACACCCTTTGGCACTGAGATCTGCAAGTT 509
         || ||||  |  || ||| || ||  ||||  |||| || || |||| | ||| | ||
Db   243 ACTTATCTTATAACTAACGCCAGAGGGGACCACCTGCCCTTCGGAGACCTCGCCTTT 302

Qy   510 CGTCGGCGCTTTCTTTTCTATTGGAACCTCTACTGCAGTGTCCTTTTCTCCACCTGCATCAG 569
         ||||  |||   |||  |  |||| | |  |||||| ||  | |||||||||| ||  | ||
Db   303 TGTACGGTTCCTCTTCTATGCCANCTACACGACGCCATCCTTCCTCCACCTGCTCATTAG 362

Qy   570 CGTGCACCGCTACATCTCCCACCACTTCGGGCGACTACGCTGGGGCGCCCCCTCG 629
         ||  | | |||||  |||||||| ||  ||| || |  |||||||  || |||||
Db   363 CTTCCAGGGCTACCTGCCACCCACTTTCCTGCCACAAGCGTGGGAGGTCG 422

Qy   630 CCTCGCAG---GCCTTCTCTCCGCCAGTTGGTTGGTGTAGCCGGCGGCCTCGTGCC 686
         ||         || | |  | | | || ||| |||||  |  ||  ||  | || |
Db   423 CCGTGCTTCTTGGTAGTGTTGGAGTCGTTGGGGAGCAGCCGCAGGTGCCTGTGGGC 482

Qy   687 CAACCTGTTCTTTGTCACAACAACAAAGGGACCACCGTTCCTTGCCCCATACACCAC 746
         |   ||||||| ||  |  |  |  || | ||  |  |||| || || |||| ||
Db   483 CACGGCAGTCTTTGCTCCGGCAGAACTTCAAGCAACTGCTGCTCCCACTGTGCTACAGCGAG 542

Qy   747 TCGGCCCTGAAGAGTTTGACCAACTATGTGCACTTCAGCTTCGGGCTGCTCATGGGGCTGCTT 806
         || |||||  | | |||  ||| | |||  || ||  ||||||||| | |||  |  | ||
Db   543 CCCACCANCCTGTCTACTTGCGCTACCCTAGGGTTATGGGGGTCATGGGGTCTCA 602

Qy   807 TGGCGGATGCAACGGCCCTGCCTGGTCACTCTGTTTGCTATGGACTCGATCGCCGGCCTGTGTCG 866
         |||| |||||| ||||| | | ||| ||  |  ||  |  ||||| | | ||  | ||||| | |
Db   603 CTTGCTGCGCCTTCAGTAGCCTTACTGGCTTGTTATTGTCCATGGCCCGCGGCCTTGTGTCG 662

Qy   867 GCC------CTTGCCAGGCGTCTGCGACAGTGCGTTCTCCGGCCGCTCCCGCTTCTCGGCCACCAT 920
         |||        ||||  |  ||| ||| | | |  ||  | ||    | | | ||||  |||| |
Db   663 CCAGGATGGCCCACGAGACCGGTGCGTGGGCCCCAAGGAGCGGCGGGCTCTTGTATGGC 722

Qy   921 AGCGTGTGGGTGCCTGACTGTCTTTTCGCTCGTTGGCTTCGTCGGCCTTTCCACATCACCCCACCAT 980
         ||| |||||| || | | |  ||| |||  ||  | || | |  | || | |  ||| | |  |||
Db   723 TGTGGTGGGTGGCAGCGTGCGTCGTTTGGCCATCAGCCTTCCTGCCCGGCCTTTTCCACATCACCAAGAGC 782

Qy   981 TTACTACCTGGCCAGGCGTGTTGGAAGCTGAC----TGCCCAGTAGTACTGAACATTGTCAAGCGT 1037
         ||    | | ||  |  | ||  |  |  | |      || | | | | ||| | | | || || |
Db   783 CTACTTGGCTGCGTCCCACGCCCGGGTGTCTCTTCTGCTGCTGGAGACCTTCGCTGC 842

Qy  1038 GGTCTATAAAGTGACTCGGCGCCCTGGGCCAGTGCCAAGCAGCTGGGATCCTGTGTGCTCTA 1097
         || | ||||  |||  || |   |  ||| ||| | |||  |   |||| | | |||  |
Db   843 TGCCTACAAAAGGCACTCGCCGCCTTCGCCAGTGGTCAACACAGTGTTCTGGACCCCAATTCTCTT 902

Qy  1098 CTTGCTCACTGGGGACAAATATCGACGTCAGCCTCCGTCAGCTCT 1141
         || ||| ||||  || | ||   || | |||  || | | ||  ||
Db   903 CTACTTCACACAACAGAAGTTCCGCGCGGCAACCCCACGATCTCT 946
```

```
RESULT  5
T47730
ID  T47730 standard; cDNA; 984 BP.
AC  T47730;
DT  01-SEP-1997 (first entry)
DE  Human placenta purinergic P-2u receptor, PNR, cDNA.
PN  WO9638558-A2.
PD  05-DEC-1996.
PF  03-JUN-1996; U08481.
PR  02-JUN-1995; US-459046.
PA  (INCY-) INCYTE PHARM INC.
PI  Au-Young J, Coleman R, Stuart SG;
DR  WPI; 97-042714/04.
DR  P-PSDB; W09433.
PT  New isolated human placenta purinergic P-2u receptor (PNR) gene -
PT  used to develop prods. for the diagnosis and treatment of conditions
PT  associated with altered expression of the receptor e.g. inflammation
PS  Claim 2; Page 27; 36pp; English.
CC  T47730 encodes a human placenta-derived purinergic P-2u receptor
CC  designated PNR. P-2u is specifically expressed in cells active in
CC  immunity. An assay for up-regulated expression of PNR can accelerate
CC  diagnosis and proper treatment of conditions caused by abnormal signal
CC  transduction due to systemic and local infections, traumatic and
CC  other tissue damage, hereditary or environmental diseases associated
CC  with hypertension, carcinomas, cystic fibrosis and other physiological
CC  or pathological problems.
SQ  Sequence 984 BP; 157 A; 360 C; 259 G; 208 T;

Query Match      14.7%; Score 210.6; DB 1; Length 984;
Best Local Similarity  54.6%; Pred. No. 8.4e-49;
Matches 489; Conservative  0; Mismatches 394; Indels 12; Gaps  3;

Qy   258 CTGTTGGTTGATGAGGATTTCAAGTTCATCCTGCTGCCTGTGAGCTATGCAGTTGTCTTT 317
         ||||     |  | | |||||||||||||||||||||||||||| ||||||||||||||
Db    51 CTGTGTCTACCGGCGGAGAACTTCAAGCAACTGCTGCTCCCACTGTGCTACAGCGAGTTGTT 110

Qy   318 TGTCTGGGGCTTGGGCCTTAACGCCCCAACCCTATGGCTCTTCATCTTCCGCCTCCGACC 377
         ||| || |||  |  | ||   ||||  | ||| || |  || |||  || ||  |||
Db   111 GGGCCGCCTGCCCTCCCGGTGAACATCTGTGTCATTACCGATCTGCAGTGCCCGGGGC 170

Qy   378 CTGGGATGCAACGGCCCACTACATGTTCCACCTGCGCATTGTCAGACACCTTGTATGTGCT 437
```

```
Db   171 CCTGACCCGCACGGCCGTGTACACCGTAAACCTTAAACCTTGCTCGTGCCTGACCTGTATATGCCTG
           ||||||                |||||||         ||  |||||
Qy   438 GTCGCGTGCCCACCCTCATCTACTATTATTGCAGCCCACGAGTGCCCCTTTGGCGTTGTCTT        230
Db   231 CTCCCCTGCCCTGTCATCTCAACATATGCCCAAGGTGATCACTGCCCTTTGGCGACTT            497

Qy   498 GATCTGCAAGTTCGTCCGGCTTTCTTTCTTCTATTGGAAGCCTTACGTACGTCCTTTTCCT        557
           |||||  |||||  ||||||      ||||||| |||||||        ||||||
Db   291 CGCCTGCCTGGTCCGCCTTCCTCTTCTATGCCAACCTGCACGGGAGGATCCTCTTCCT          350

Qy   558 CACCGTGGGCCATCAAGCGTGCCACCGTACCTGGGCAATCTGCCACACTTGGGCACTACCGC--    616
Db   351 CACCTGCATCAGCTTCCAGCCGTACCTGGGCATTCTGCCACACCCGGCCCCCTGGGCACAA       410

Qy   675 CTGCCTCGTGCCCAACCTGTTTGTGTCACAACCAGAACAAAAGGGACCACCGTCCTGTG        734
Db   471 CCAGTGCCTGCCACACAGTTCCGCTGCCACAGGCGTAACCGCACCGCGTCTG              530

Qy   735 CCATGACACCACTCGGCCGTGAAGAGTTTGACCACTATGTGCACTTCAGCTCGGCGTCAT        794
Db   531 TTATGACCTCAGCCCGGCGCTGCCCTGGCACCGACCACTATATGGGGAT            590

Qy   795 GGGGCGTGCTTGCTTTTGGCGGTGCCCTGCCTGGTCAACAGCTGCTGGA              854
Db   591 TGTCATCGGCTTCCTGCGTGCCCTTTGCTGCCCTTTGCTGGGTT                    650

Qy   855 TCGGCCTGTATCAGGCC---------CTTGCCAGGCGTCTCGCACAGTGCTATGCAGTTGTCTT    908
Db   651 CCGCCGTGCGCCAGGATGGCCCGGAGAGAACTTCAAAGCAACTGTCTCCCACCCTGTGTATTCGGCGGTGCT  710

Qy   909 TCTCCCGCCACCATAGCGTGTGGTGTCGCTGACTGTCTTTGCTGTTCGTTGCCTTCGTGCCACAT  968
Db   711 GGGCCCGGCCATGGGCGGTGGGGTGTTGGGCATCAGCTTCCTGTCTTCTCTGCCTTTCACAT      770

Qy   969 CACCCGCCACCATTTAC----TACCTGGCCAGCGCTGTTGGAAGCGTGACTGCCGAGTACTGAA   1025
Db   771 CACCAAGACACACCTACCTGGCCGACGTGCCCTCGACGCCGGGCGCCCTGCCACTGTATTGGA       830

Qy  1026 CATTGTCAACGTGGTCTGTATAAAAGTGACTCGGCCCCCTGGGCCCTGGGCCACTGCCTGGA     1085
Db   831 GGGCCTTTGCAGCGGCCTACAAAAGGCACGGCGGCCGTTTGCCAGTGCCAACAGCGGTGCTGGA     890

Qy  1086 TCCTGTGCTCTACTTCCTCACTGGGGACAAAATATCGACGTCAGCTCGGTCAGCTC   1140
Db   891 CCCAATCCTCTTCTACTTCACCCAGAGAGAAGTTCCCCGGGGACCACATGAGCTC    945
```

```
RESULT   6

T70000

ID   T70000 standard; cDNA; 984 BP.
AC   T70000;
DT   26-JAN-1998 (first entry)
DE   P2U purinergic receptor gene homologue.
KW   primer; PCR; extend; partial cDNA; human cathepsin B; sequence analysis;
KW   complete gene; detection; amplification; 5'-untranslated region;
OS   Homo sapiens.
PN   WO9638591-A1.
PD   05-DEC-1996.
PF   05-DEC-1996;
PR   03-JUN-1996;   U08501.
PR   01-DEC-1995;   US-566334.
PR   02-JUN-1995;   US-459046.
PR   05-JUN-1995;   US-462355.
PR   07-JUN-1995;   US-487112.
PR   15-NOV-1995;   US-006809.
PA   (INCY-) INCYTE PHARM INC.
PI   Guegler KJ;
DR   WPI; 97-034397/03.
```

```
CC   This is the full length cDNA of a novel P2U purinergic receptor gene
CC   homologue. It was identified using the inventive method and comparison
CC   between the partial sequence found in Incyte clone 179696 (from a
CC   placental cDNA library) and the GenBank sequence, accession number
CC   RNU09402, a G-protein coupled surface receptor from rat. The method
CC   comprises: (a) subjecting nucleic acid from a genomic or cDNA library
CC   expected to contain partial cDNA, to reaction with a PCR primer pair,
CC   which is able to anneal to opposite strands of the template DNA and to
CC   initiate outward DNA synthesis, where one is extended by DNA polymerase
CC   in the antisense direction and the other in the sense direction, to
CC   generate PCR products; and (b) purifying the products, and identifying
CC   extended sequences derived from the partial cDNA or genomic DNA. The
CC   method is used to produce much longer DNA (complete genes in some cases)
CC   from libraries which have been used for sequencing. It may also be used
CC   to extend known genomic sequences for detection and amplification of
CC   5'-untranslated, and/or promoter sequences. Up to 96 genes may be
CC   processed at a time, and the extended or complete sequence is obtained
CC   within 6-10 days, an approximate 10-fold improvement in productivity.
SQ   Sequence  984 BP;  157 A;  360 C;  259 G;  208 T;
```

```
Query Match       14.7%;   Score 210.6;  DB 1;  Length 984;
Best Local Similarity  54.6%;   Pred. No. 8.4e-49;
Matches  489;   Conservative  0;  Mismatches 394;  Indels  12;  Gaps  3;
```

```
Qy   258 CTGTTGGTTTGATGAGGAGATTTCAAGTTCATCCGTGCCTGTGGAGCTATGCAGTTGTCTT    317
           ||||                  ||  |  ||||||||      ||||| ||| ||
Db    51 CTGTGTCTACCGCGGAGAACTTCAAAGCAACTGTCTCCCACCCTGTGTATTCGGCGGTGCT    110

Qy   318 TGTGTCTGGGCTTGGGCCTTAACGCCCCAAACCCTATGGCTCTTCATTCTTCCGCCGACC    377
           ||||                   |||                ||||
Db   111 GGGCCCGGCCATGGGCGGTGGGGTGTTGGGCATCAGCTTCCTGTCTTCTCTGCCTTTCACAT  170

Qy   378 CTGGGATCGCAACGGGCCACCTACATCAATGTTCCACCTGGCATTGTCAGGACACCTTGTATGTGCT  437
           |||||||                  ||||||||| |      |||||  |||
Db   171 CCTGACCCGCACGGCCGTGTACACCGTAAACCTTAAACCTTGCTCGTGCCTGACCTGTATATGCCTG   230

Qy   438 GTCGCGTGCCCACCCTCATCTACTATTATTGCAGCCCACACAACCACTGCCCTTTGGCGACTGCACTGA  497
           ||||                   |||||||        |||||||||||
Db   231 CTCCCCTGCCCTGTCATCTCAACATATGCCCAAGGTGATCACTGCCCTTTGGCGACTT    290

Qy   498 GATCTGCAAGTTCGTCCGGCTTTCTTTCTTCTATTGGAAGCCTTACGTACGTCCTTTTCCT    557
           |||||  |||||  ||||||      ||||||| |||||||        ||||||
Db   291 CGCCTGCCTGGTCCGCCTTCCTCTTCTATGCCAACCTGCACGGGAGGATCCTCTTCCT    350

Qy   558 CACCGTGGGCCATCAAGCGTGCCACCGTACCTGGGCAATCTGCCACACTTGGGCACTACCGC--  616
           ||||||||||                       ||| |||||||||||||
Db   351 CACCTGCATCAGCTTCCAGCCGTACCTGGGCATTCTGCCACACCCGGCCCCCTGGGCACAA     410

Qy   616 -TGGGCGGCGGCCCTCGCGCCAGGCCGTTCTCGTGCCAGGTTTGGTTGGTTCGTGAGCCGG     674
           ||||||||||||    ||   ||||||||| || ||  ||||| | ||||||
Db   531 TTATGACCTCAGCCCGGCGCTGCCCTGGCACCGACCACTATATGGGGAT             590

Qy   735 CCATGACACCACTCGGCCGTGAAGAGTTTGACCACTATGTGCACTTCAGCTCGGCGTCAT    794
           ||||                     ||| |||||||||||
Db   471 CCAGTGCCTGCCACACAGTTCCGCTGCCACAGGCGTAACCGCACCGCGTCTG          530

Qy   675 CTGCCTCGTGCCCAACCTGTTTGTGTCACAACCAGAACAAAAGGGACCACCGTCCTGTG    734
           ||||                    |  |||||
Db   411 ACGTGGGGGGCCCCGGGCTGGCCCTGCCCTGGTCAGTGCCTGCTTAGTGGGACTCATGGCCTCG   470

Qy   795 GGGGCGTGCTTGCTTTTGGCGGTGCCCTGCCTGGTCAACAGCTGCTGGA              854
           |||||                    |||          |||||
Db   591 TGTCATCGGCTTCCTGCGTGCCCTTTGCTGCCCTGCTCGGTTGGCGTCTCTCACTGGTCCCTCGC   650

Qy   855 TCGGCCTGTATCAGGCC---------CTTGCCAGGCGTCTCGCACAGTGCTATGCAGTTGTCTT    908
           ||||||                    |||  |||||||
Db   651 CCGCCGTGCGCCAGGATGGCCCGGAGAGAGCCTGTGCGGGCCAGGAGAGCCTGTGGCCGCAAGGC  710
```

```
Qy    909  TCTCCGCACCATAGCTGTGTGCTGTGACTGTGTCTTTTGCTTCGTTCGTTCCTTTCCACAT  968
           |||||        |||||||||||||     ||||||||||||  |||||||| |||||||
Db    711  GGCCCGCATAGGCCCTGGGCGTGGGTGGCTGCTGTCTTTGGCATCAGCTTCTTCCTTTCACAT  770

Qy    969  CACCCGCACCATTTAC---TACCTGGCCAGGCTGTTGGAAGCTGACAGCTTCTTCCTTTTTT  1025
           ||||||||||||||     |||||  ||||||||||||||||||||||||||||||
Db    771  CACCCAAGACAGACCTTACCTGGCCAGGCTGTTGTGGAAGCTGACAGCTTCTTCCTTTTTT  830

Qy   1026  CATTGTCAACGTGGTCTATAAAAGTGACTGGCGCCCCCTGGCCAGTGCCAGTGCCCTGGA  1085
           |||||||| ||| |||| || |||||||||
Db    831  GGCCTTTGCAGCGGCCTACAAAAGGCACGCGCCCGGTTGCCAGTGCCAACAGCGTGCCTGGA  890

Qy   1086  TCCTTGTGCTGCTTACTTTGCTCACTGGGGACAAAATATCGACGTCAGCTCAGCTGC  1140
           |  ||| ||  |||| ||||||||||| ||||||||||||
Db    891  CCCATCCTTCTTCTTACTTCACCCAGAGAAGAAGTTCCGCGCGGCGACCACATGAGCTC  945

Qy   RESULT    7
Db   Q66178

ID   Q66178 standard; DNA; 1901 BP.
AC   Q66178;
DT   07-FEB-1995 (first entry)
DE   Seven transmembrane receptor (R12) coding sequence.
KW   Primer; seven transmembrane receptor; receptor; amplification; ss.
OS   Homo sapiens.
FH   Key    Location/Qualifiers
FT   cds    701..1720
FT          /tag= a
FT          /product= Seven transmembrane receptor (R12).
PN   WO9412635-A.
PD   09-JUN-1994.
PF   17-NOV-1993; U11153.
PR   17-NOV-1992; US-977452.
PA   (ICOS-) ICOS CORP.
PI   Godiska R, Gray PW, Schweickart VL;
DR   WPI; 94-200264/24.
DR   P-PSDB; R53753.

CC   DNA encoding seven trans:membrane receptors - used to develop
CC   prods. for use as therapeutic or diagnostic agents for conditions
CC   involving the receptors.
PS   Example 10; Page 79-80; 100pp; English.
PT   PCR using two primers (Q66174, Q66175) was performed to amplify a
PT   partial sequence of the R20 seven transmembrane coding sequence
PT   which was later used as a probe for isolating the R20 genomic clone
PT   of the R20 gene, from a human placenta DNA library. During the isolation
CC   which had significant homology to other seven transmembrane
CC   receptors.
CC       The probe was used to screen a human genomic foetal
CC   liver DNA library, and while the R20 gene could not be identified in
CC   this library, several weakly hybridising clones were plaque
CC   purified, subcloned and sequenced. The two clones were designated
CC   R2 (Q66177) and R12. This is the coding sequence of the R12 clone.
SQ   Sequence 1901 BP; 381 A; 640 C; 508 G; 372 T;

Qy    255  GGACTGTTGGTTTGATGAGGATTTCAAGTTCATCCTCGTGAGCTATGCAGTTGT  314
           |||||||       |||||||| |||||||||||||||||||||
Db    763  GCAATGTGGCCAGGAGAGCCACTGGAGAAACATGCTGTTGCCTCCTTCTACCTTCTGGA  822

Qy    315  CTTTGTCCTGGGGCTTGGGCCTTAACGCCCCAACCTATGGCTCTTCAATCTTCCGCCTCCG  374
           ||||| ||||||||| |||      |||||||||||  |||||||||||
Db    823  TTTTATCCTGGCTTTAGTTGGCAAATACCCTGCCTCTGGGCTTTTCATCCGAGACCACAA  882

Qy    375  ACCTGGGATGCAAACGGCCACCTACATGTTCCACACCTGTATGT  434
           ||||      ||||||||||||||||  |||||||||||||
Db    883  GTCCGGGACCCCGGCCAACGTGTTCCTGAATGCATCTGGCCGCCGACTTGTCGTGCGT  942

Qy    435  GCTGTCGCTGCCCACCCTCATCTACTATTATCCAGCCCACAACCACTGCCCTTTGGGCAC  494
```

Query Match	9.0%;	Score 128.4;	DB 1; Length 1901;
Best Local Similarity	48.9%;	Pred. No. 5.4e-26;	
Matches 447;	Conservative 0;	Mismatches 446;	Indels 21; Gaps 3;

```
Db    943  GCTGGTTCCTTGCCCACCACCGGTCGTCTACCAACTTCTCTGGAACCACTGGCCATTTGGGGA  1002
           ||||||||||||||||     |||||||||   ||||||||||||| ||||||
Qy    495  TGAGAATCTGCAAGTTCGTTCCGCCTTTCTTTTCTTTCCTTTATTGGAAGCTTCTACTGCAGTGTCCTTTT  554

Db   1003  AATCCCATGCCGTCTCACCGGCCTTCCTCTTTCTTACCTCAACATGTACGCCAGCATCTACTT  1062
           ||||||||||||| |||  |||||||||| ||||||||||||
Qy    555  CCTCACCTGCCATCAGCGTGCACCCGCTACCTGGGGATCGCCATCCTGCCACACCGAC  614

Db   1063  CCTCACCTGCATCAGCGGCGACGTTTCCTTGCCCATTGTGCACCGGTCAAGTCCCTCAA  1122
           |||||||||||||||| |||||||||||||||||||||||||||
Qy    615  CTGGGGCGCCCGCCCTGCCCTGCCAGCGCGGCCTTCTCTGCCCCAGTTGTGTCGGCGGT  674

Db   1123  GCTCCGGCCAGGCCCCTCTTACCACACCCACTGGCCCTGTGCCCTGTGGGGGTGGGGCTGT  1182
           |||||| |||||||||||| ||| |||||| ||| ||||||||||||
Qy    675  CTGCCTCGTGCCCAACCTTGTTCTTTGTCACACAGGCGAACACAAAAGGGAACCACACGGTGGTCTG  734

Db   1183  GGCCATTGGGCCCCGCTGCCTGGTGAGCCCACAGACCGTGCAGACCAAACCACACGTGTCAT  1242
           ||||      |||||    ||||||||| |||||||||||
Qy    735  CCATGACCACCTCGGCCTGAAAGAGTTGACCACTATGTGCACTTCAGCTCGGCGGGCAT  794

Db   1243  CCTGCCAGCTGTACCGGGGAGAAGGCCTCCAACA  1290
           ||||  ||||||||   ||||||||||||          ||| |||||||||
Qy    795  GGGGCTGCTCTTGCGGTGCCCTGCCTCCTGTCACTCTTGTTTTGCTATGGAGACTCCATGGCTCG  854

Db   1291  GGCAGTGGCCTTCACCTTCCGGTTCATCACCAGGTCACCTGCTGTACCTTGCTACCTCCGTGATCATCG  1350
           ||||  |||||||||||||||||||||||||||||||||
Qy    855  TCGCCTGTATCAGCCTTGCCCAGGCTTCCCACACGTCGTCTTCGCGCCTCCGGCGTCTCCG  914

Db   1351  CAGCCTGCCGGCCAGGGCCTCG-----TGTGGAGAAGGCCACCACCTGCAAGACCAAAGGGCAGTGCG  1404
           |||||||||||||||           ||||  ||| ||||
Qy    915  CACCAGTAGCCTGGTGGTGCTGACTGTCTTTGCCTGCTGCTTCGCTTCCACCGTCCACCGCCTTCAGCTCGGGCTCA  974

Db   1405  CCTGCCAGCTGTACCGGGGAGAAGGCCTCCAACA  1464
           ||||||||||||||       ||| |||||||||
Qy    975  CACCATTTACTTACCTGGCCAGGCTGCTGTTGGAAGCTGACTGCCGAG---TACTGAACATTGT  1031

Db   1465  CTCCGTCTACGTGCGTGCACTACCGGCCAGCCATGGGCCTCCTGGGCAGTGGCCATCCT  1524
           ||||   |||||||||||||||||||||
Qy   1032  CAACGTGGTCTATAAAAGTGACTCGGCGCCCCAGGTGCCCAGTGCCACCAGGGGATCCTGT  1091

Db   1525  GGCCCTTGGCAAACCGCATCACCTCCTGTGGCCCCCTGGCCAGTGCCACCAGGGGCACCCCAT  1584
           ||||||||| ||||||||||||||||||||
Qy   1092  GCTCTACTTGCTCACTGGGGACAAAATATCGACGTCAGCGCTCAGCGTCTGTGGTGG  1151

Db   1585  CATGTATTTCTTCCGTGGGCTGAGAAGTTCCGACCACGGCGGTTGCTCTGTGGCAA  1644
           ||||||   |||  |||||||||||
Qy   1152  CAAGCCCCAGCCCC  1165
           |||||||||
Db   1645  AAGGCTCAAGGGCC  1658

RESULT    8
ID   T44092 standard; cDNA; 2453 BP.
AC   T44092;
DT   28-FEB-1997 (first entry)
DE   Human G-protein thrombin-like receptor cDNA.
KW   G-protein; thrombin; receptor; diagnosis; gene therapy;
KW   haemophilia; wound healing; restenosis; angina; inflammation; ss.
OS   Homo sapiens.
FH   Key    Location/Qualifiers
FT   cds    548..1567
FT          /tag= a
FT          complement (548..565)
FT   primer_bind  /note= "5' primer binding site"
FT          /tag= b
FT          1547..1564
FT   primer_bind  /tag= c
FT          /note= "3' primer binding site"
PN   WO9639438-A1.
PD   12-DEC-1996.
PF   06-JUN-1995; U07180.
```

```
PR   06-JUN-1995; WO-U07180.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Gocayne JD, Li Y, Ruben SM;
DR   WPI; 97-043073/04.
DR   P-PSDB; W07617.
PT   Human G-protein thrombin receptor, HIBEB69 - useful to identify
PT   (antagonists, for treatment of angina, restenosis, wound healing
PT   etc.
PS   Claim 1; Fig 1A-D; 58pp; English.
PA   A cDNA clone (T44092) codes for a human 7-transmembrane receptor
CC   (W07617) putatively identified as a G-protein thrombin receptor.  It
CC   was discovered in a human infant brain cDNA library. G-protein
CC   thrombin-like receptor polynucleotides can be used to produce
CC   recombinant receptor in host cells (e.g. E. coli, COS-7, Sf9), to
CC   design probes for detecting diseases related to receptor gene
CC   mutation, and in gene therapy to treat haemophilia or promote wound
CC   healing; antisense sequences can be used to treat inflammation,
CC   restenosis, angina, etc.
SQ   Sequence  2453 BP;
```

```
Query Match             9.0%;   Score 128.4;  DB 1;  Length 2453;
Best Local Similarity   48.9%;   Pred. No. 6e-26;
Matches  447;  Conservative    0;  Mismatches  446;  Indels   21;  Gaps   3;
```

```
Qy    255 GGACTGTGTTGGTTGATGAGGAATTTCAAGTTCATCCTGCCTGTGAGCTATGCAGTTGT 314
          |||  ||| | ||| |   ||| ||||||  |  ||| |  |  | ||| | || ||| |
Db    610 GCAAGTGTGGCCAGGAGACGCCACTGAGACAATGCTGTTCTCGCCTCCTTCACCTTCTGGA 669

Qy    315 CTTTGTGCTGGGCCTTGGGCCTTAACGCCCCAAACCCTATAAAGTGCTTCTTCATCTTTCCGCCGTCTTCCG 374
          |||| |||| | |||| ||| || | | | | |  ||   |||| | | |  |  |  |||  |||
Db    670 TTTTATCCTGCCTTTAGTTGGCAAATACCCTGGCTTTCATCCGAGACCACAA 729

Qy    375 ACCCTGGGATGCAACGGCCACCTACAGTGTTCCACCTGGCATTGTCAGACACACTTGTTATGT 434
          |||      ||| | | | || | | | ||||  ||| |||||||  ||| | ||||
Db    730 GTCCGGGAACCCGGCCAACGTGTTCCTGATCCATCTGGCCGTGGCCGACTTGTCGTGGGT 789

Qy    435 GCTGTGCGCTGCCACCCTCATCTACTACTATTATGCAGGCCACAACCACTGGCCCTTTGGCAC 494
          |||| |||| |||| |||| ||| || || | || || || ||| | |||||  |||  |||
Db    790 GCTGGTTCCTGCCACCCGCTGGGTACCACCTTCTCTCGGAACCACACTGGCCCATTGGGGA 849

Qy    495 TGAGATCTGCAAGTTCGTCCGCCTTTCTTTTCTTATTGGAACCTCTACTGCAGTGTCCTTTT 554
          |||     ||| |  | || | |||  ||| |||||| || | || | |||| | |||
Db    850 AATCGCATGCCGTCTCACCGGCTTCCTCTTCTTCACCCTCAACATGTACGCACCATCTACTT 909

Qy    555 CCTCACCTCCATCAGCGTCACCGCCGCCACCGTGGCCGCATCTGCCACCACTTCGGCCACTACG 614
          ||| |||| | | |||   | | | | | |||| ||| |  || || |  || |||| | |||
Db    910 CCTCACCTGCATCACCGCCACGGTTCCTGCCAATGTGACACCGTCAAGTCCCTCAA 969

Qy    615 CTGGGCGCCCCGCCCTGCCCCGCCAGGCGCCTTCTGCTGGCAGTTTGGTTGGTTGGTAGCCGG 674
          ||| | ||| || |||| ||| |||| | |||||  ||||| | ||||  || || |||||
Db    970 GCTCCGCAGGCCCCTTCTACCGCACCACTGGGCCTTCCTGCTGTGGGCTGGGTGGTGCCGT 1029

Qy    675 CTGCCTCGTCTGCCCAACCTGTTTCTTTGTCACAACCACAGACAACAAAGGGACCACCGTCCTGTG 734
          |||  | | | |||  ||| ||| | || | |||  ||| ||| ||| ||||| |||| | |
Db   1030 GGCCATGGCCCCCGTGCTGCTGTGAGCCCACAGACCTGGTGCTCG 1089

Qy    735 CCATGACACCACTCGGCGCTCGGCCTGAAGAGTTGACCACTATGTGCACTTCAGCTGGGCGTCAT 794
          || || | | | ||| ||  ||| |||| || | | ||| |  ||| | | |||| || |||
Db   1090 CCTGCAGCGTGTACCGGGAGAAGGGCTCCCACCA---------TGCCCTGGTGTCCCT 1137

Qy    795 GGGCCTGCTCTTTGCGTGCCCCTGCTCACTCTGTTGTTTGCTATGGACTCATGGCTCTG 854
          ||  ||| || |  | ||  ||| |  | ||| | | ||| || | || | ||  | |
Db   1138 GGCAGTGGCCTTCACCTTCCCGTTCATCACCAGCTACCTGCTGATCATCCG 1197

Qy    855 TCGGCCTGTATCAGAGCCCTTGGCGCTGCTGCACAGTGCGTCTTCTCGGCGCTCGGCCTCTCGCCG 914
          ||| || | | ||| ||| | || | | ||| || || ||  |  | |||  |||  |
Db   1198 CAGGCCTGCGCCAGGCCTCCG------TGTGGAGAAGCCCTCAAGACCAAGGCAGTGCG 1251

Qy    915 CACCATAGCTGTGGGTGCTGACTGTGTCTTGTGCTTGTGCCTTTCCACATCACCCG 974
          || ||   | ||| |||   |  |||||||||  ||| ||| || |||||| | |||
Db   1252 CATGAATCGCCATAGTGCTGGCCCATCTCCTGGGTCGTCGCCCTACCACGTCAAACCG 1311
```

```
RESULT  9
ID   V18356
AC   V18356;
DT   25-SEP-1998 (first entry)
DE   Human R12 seven transmembrane (7TM) receptor genomic DNA.
KW   V28; placenta; seven transmembrane receptor; 7TM; signal transduction;
KW   immunology; inflammation; R12; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             701..1720
FT                   /tag= a
FT                   /product= "Human R12 seven transmembrane receptor"
```

```
Qy    255 GGACTGTGTTGGTTGATGAGGAATTTCAAGTTCATCCTGCCTGTGAGCTATGCAGTTGT 314
          |||  ||| | ||| |   ||| ||||||  |  ||| |  |  | ||| | || ||| |
Db    763 GCAAGTGTGGCCAGGAGACGCCACTGAGACAATGCTGTTCTCGCCTCCTTCACCTTCTGGA 822

Qy    315 CTTTGTGCTGGGCCTTGGGCCTTAACGCCCCAAACCCTATAAAGTGCTTCTTCATCTTTCCGCCGTCTTCCG 374
          |||| |||| | |||| ||| || | | | | |  ||   |||| | | |  |  |  |||  |||
Db    823 TTTTATCCTGCCTTTAGTTGGCAAATACCCTGGCTTTCATCCGAGACCACAA 374

Qy    375 ACCCTGGGATGCAACGGCCACCTACAGTGTTCCACCTGGCATTGTCAGACACACTTGTTATGT 434
          |||      ||| | | | || | | | ||||  ||| |||||||  ||| | ||||
Db    883 GTCCGGGACCCCGGCCAACGTGTTCCTGATGGCCATCTGGCCGTGGGC 883

Qy    435 GCTGTGCGCTGCCACCCTCATCTACTACTATTATGCAGCCCAC
```

(Result 9 block continues)

```
P-FSDB; W48733.
DR   Goliska R, Gray PW, Schweickart VL;
WPI; 98-332132/29.
```

The present novel sequence represents the human R12 genomic DNA isolated from a human genomic fetal liver DNA library. The invention claims for the full length V28 genomic DNA sequence (V18343) isolated from a human placenta genomic library. The V28 (W48722) and R12 proteins are seven transmembrane (7TM) receptors which are probably involved in signal transduction. The invention also claims that cells transformed with V28 DNA can be used to produce the recombinant polypeptide, to produce anti-V28 antibodies or in screening assays for V28 agonists or antagonists. The antibodies, agonists and antagonists could then be used to modulate V28 receptor-ligand binding, for e.g. in immunological and/or inflammatory events in vivo.

SQ   Sequence 1901 BP;

```
Query Match             9.0%;   Score 128.4;  DB 1;  Length 1901;
Best Local Similarity   48.9%;   Pred. No. 5.4e-26;
Matches  447;  Conservative    0;  Mismatches  446;  Indels   21;  Gaps   3;
```

```
Db   943  GCTGGTCGTGCCCACCCGCCTGGTCTACCACTTCTCTGGGACCACACTGGCCATTTGGGGA 1002
              ||||||||||| | |||||||| ||||||| |   ||||||||||| |
Qy   495  TGAGATCTGCAAGTTCGTCCGGCTTTCTTTCTATTGGAACCTTCTACGTGTCCTTTT 554

Db  1003  AATCGGCATGCCGTCTCACCGGCTTCTCCTTCTTCTTTCTTTACCTCAACATGTACGCCAGCAATCTACTT 1062
Qy   495
Db  1063  CCTCACCTGCATCAGCGCCGACCGGTTTCCTGGCCATTGTGCACCGGTCAAGTCCCTCAA 1122
Qy   555  CCTCACCTGCATCAGCGTGCACCGCTACCTGGGCATCTGCCACCACTTCGGCGACTACG 614
Db  1123  GCTCCGCGAGGCCCCTCTACGCACACCTGGCCTTCTTCCTGTGGGTGGGTGGCGGTGTT 1182
Qy   615  CTGGGGCGCCGGCCCCTCGCCTCGCACGGCCTTCTCTGCCTGCCAGTTTGGTTGGTTGCTCGTTAGCCGG 674
Db  1183  GGCCATGGCCCGCCGCCTGCTGGTGAGCCCACAGACCTGGCACCAACCACCACCGTCCTGTG 1242
Qy   675  CTGGCCTCGTCGTCCCAACCTGTTCTTTGTCACACCAGCAACAAAAGGACCACCGTCCTGTG 734
Db  1243  CCTGCAGCGTGTACCGGGAGAAGGGCTCCCACCA------------TGCCCTGGTGTGTCCCT 1290
Qy   735  CCATGACAACCACTCGGCGGCCTGAGAGTTTGACCACTATGTGCACTTCGGCGTGCATCAT 794
Db  1291  GGCAGTGGCCTTCACCTTCCCGTTCATCACCACGTGCATTCCCGCCCTCCGACCCGTGGGAATGCAAGCGCCACC 1350
Qy   795  GGGGGCTGCTCGTTCTTTGGGCGTCCTGGCCTGGTCTCTTGTTGCTATGGGACTCATGGCTCG 854
Db  1351  CAGCCTGCGGCAGGGCCTGCG-------TGTGGAGGAAGCGCCCTCAAGACCAAGGCCAGTGCG 1404
Qy   855  TCGCCCTGGTATCAGCCCTTGCCAGGCTCTGCACAGTCGTCTTCTCGCCTCCGCTCTCTCCG 914
Db  1405  CATGACAGCGCCATAGTGCTTGGCGTCTGCTGTTTGCTGTGGCCCTACCACGTCAACACG 1464
Qy   915  CACCATAGCGTGTGGTGCTGACTCTTCTTTGCTGTCGCTGCCTTTCCACATCACCCG 974
Db  1465  CTCGGTCTTACGTGCTGCACTACCGGCCATGGGGCCTTCCTGGGCTTCCTTCCTGGGATTCCTGT 1524
Qy   975  CACCATTTACTTACCTGGGCCAGGCTCGTTGGGAAGCTGACTGACTGCCCGAG---TACTGAACAT TGT 1031
Db  1525  GGCCCTGGCAAACCGCATCACCTCCTGCCTCAAGGGGCAGCTCGACCCCAT 1584
Qy  1032  CAACGTGGGTCTATAAAGTGACTCGGCCCCGGCCAGTGGCCAACAGCTGCCTGGGGATTCCTGT 1091
Db  1585  CATGTATTTCTTCGTGGCTGAGAAGTTCGGCACCAGCCCTGTGCAAACTTGCTCTGTGTGGCAA 1644
Qy  1092  GCTCTACTTGCTCTCACTGGGGGACAAAATATCGCAGTCTTCAGCTCGTCCAGCTCTTGTGGGTGGTGG 1151
Db  1645  AAGGCTCAAGGGCC 1658
Qy  1152  CAAGCCCCCAGCCCC 1165
```

DR   P-PSDB; R27793.
PT   Isolated human platelet factor 4 super-family receptor
PT   polypeptide and corresp. antibodies and DNA - useful as
PT   diagnostic and screening agents, and for treating inflammation or
PT   PF4AR-mediated disorders.
PS   Claim 7; Fig 5; 78pp; English.
CC   The IL-8 receptor cDNA sequence was isolated (see Q29505) and a
CC   874bp sub-fragment of the coding sequence was used as a probe to
CC   screen human cell line HL60 and human peripheral blood lymphocyte
CC   cDNA libraries. Two new gene sequences were found that are clearly
CC   related to the IL-8 receptor. One of these was contained in clone
CC   8rr.9 and is predicted to encode an amino acid sequence which is
CC   36% and 38% identical with the high and low affinity IL-8 receptor
CC   sequences, respectively. See also Q37107.
SQ   Sequence   1679 BP;   327 A;   532 C;   457 G;   363 T;

```
Qy   277  TTCAAGTTCATCCTGCTGCCCTGAGCTATGCCAGTTGTCTCTTTGTGCTGGGGCTTGGGCCTT 336
Db   513  TTCAAGGCCGGTGTTCGTGCCCGTGGGCCTACACGCCTACACCTCCTCCTGGGGCCGGTGATCGGC 572
Qy   337  AACGCCCCAACCCTATGGCTCTTCATCTTCCGGCCTCCGACCCGTGGGAATGCAAGCGCCACC 396
Db   573  AACGTCCTGGGTGTCGTTGGTGATCCTGGAGCGCACCGGCACGCAGCAGTTCCACGGAGACC 632
Qy   397  TACATGTTCCACCTGGCCATTGTCAGACACCTTGTATGTGCTGTCGCGCTGCCACCTCATC 456
Db   633  TTCCGTGTTCCACCTGGCCGTGGCCGAACCTCCTGGTCTTCATCCTTTGCCGTG 692
Qy   457  TACTATTATGCAGCCACACCACTGGCCCTTTGGCCACTGAGAATCTGCAAGTTCGTCCGC 516
Db   693  GCCGAGGGCTCGTGTGGG------CTGGGTCCTGGGGACCTTCCTCTTGCAAAACTGTGATT 746
Qy   517  TTTCTTTTCTTCTATTGGAACCTCTACTGCCAGTGTGTCCTTTTCCTCACCTGCATCAGGGCAC 576
Db   747  GCCCTGCACAAAGTCAACTTCTACTGCCAGCAGCCGTGCTTCCTCTGGCCATGCCGTGGAC 806
Qy   577  CGCTACCTGGGGCATCTGCCACCCACTTCGGCGACTACGCTGGGGGCCGCGCCCTCGCCTCGCA 636
Db   807  CGCTACCTGGCCATTGTCCACGGCCGTCCATGCCTACCGCCACCGCGCCTCCTCTCCATC 866
Qy   637  GGCCTTCTCTGCCTGGCCAGTTTGGTTGGTTGCGTAGCCGGGCTGCCCTCGGTGCCCAACTGTTC 696
Db   867  CACATCACCCGTGTGGGACCACTCTGGGGTGGGCTTCCTCTTGCCTTGCCAGAGATTCTC 926
Qy   697  TTTGTCACACCAGCAACAAAAGGACCACCGTCCTGTGCCATGACAACCACTCGGCCTGAA 756
Db   927  TTCGCCAAAGTCAGCCAAGGCCATCACAACAAACTCCCTGCCACGTTGCACCTTCTCCCAA 986
Qy   757  GAGTTTGACCACTATGTGCACTTCAGCTCGGCGTGTCATGGGGCTTCGCTCTTTTGGGGTGCC 816
Db   987  GAGAACCAAGCAGAAAACGCATGCCTGGTTCACCTCCGACGATTCCTCTACCACAATGTGGCGTGGGA 1046
Qy   817  TGCCTGGTCACTCTTGTTTGCTATGGACTCATGGCTCGGTCGTCCGGCGCGTATCAGCCCTTGCCA 876
Db  1047  TTCCTGCTGCCCATGCCTGGTGATGGGCCTGGTGCTACGTGGGGGTAAGTGCACAGGTTGCGC 1106
Qy   877  GGCTCTGCACAGTCGTCTTCTCGCCTCCGCTCTCTCCGCACCAATAGCTGTGGTGCTGACT 936
Db  1107  CAGGCCCACGGCGCCCCTCAGCGGCAGAAGGCAGTCAGGGTGGCCATCCTTGGTGACAAGC 1166
Qy   937  GTCTTGCGTGTCTCGCTTTCGTTCCCTTTCCACATCACCCCGACCCATTTACTTACCTGGCCAGG 996
Db  1167  ATCTTCTTCCTCTCGTCACCCTACCACACGTCATCTTCCTGACACCCTGCCGAGG 1226
Qy   997  CTGTTGGAAGCTGAC-------TGCCGAGGTACTGAACATTGTCAAGGTGGTTCTATAAAGTG 1050
Db  1227  CTGAAGGCCGTGGCACAAAACCTGCACAAGCTGAAATGGCTCTCTCCCGTGCCCATCACCATG 1286
```

RESULT 10
Q37107
ID   Q37107 standard; DNA; 1679 BP.
AC   Q37107;
DT   12-MAR-1993 (first entry)
DE   New platelet factor 4 receptor superfamily member PF4AR1i.
KW   IL-8R; G-protein coupled receptor family; rhodopsin superfamily;
KW   pro-inflammatory cytokine; 8rr.9; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   cds   369..1487
FT   /*tag= a
FT   /product= PF4AR1i

RN   [1]
RC   WO9217497-A.
PN   15-OCT-1992.
PD   23-MAR-1992; U02317.
PF   29-MAR-1991; US-677211.
PR   19-DEC-1991; US-810782.
PA   (GETH) GENENTECH INC.
PI   Holmes WE, Lee J, Wood WI;
DR   WPI; 92-366191/44.

Query Match   7.58;   Score 107.6; DB 1; Length 1679;
Best Local Similarity   47.0%;
Matches 409;   Conservative 0;   Pred. No. 2.8e-20;   Indels 12; Gaps 2;
                           Mismatches 449;

```
Qy   1051  ACTCCGGCCCCTGGCCAGTGCCCCAGCTGCCGCCAAACAGCTGCTGTGCCTCTGTGCTCCTTACTTGCTCACTGGG 1110
           ||||||||||||| ||||||||| ||||||||| || ||||||  |||| |||  |||  ||| |||| ||
Db   1287  TGTGAGTTCCTGGGCCTGGCCCTGGCCCCCATGCTGCTCCAACCCATGCTCTACACACTTTCGCCGGC 1346

Qy   1111  GACAAATATCGACGTCAGCTCCGTCAGCTC 1140
           ||||||||| ||||||| |||||||||||||
Db   1347  GTGAAGTTCCCAGTGACCTGTCGCGGCTC 1376
```

RESULT 11

```
Q80522
ID   Q80522 standard; cDNA; 1679 BP.
AC   Q80522;
DT   18-JUL-1995 (first entry)
DE   Human lymphocyte PF4AR cDNA.
KW   Interleukin-8 receptor; IL-8 receptor; PF4AR;
KW   platelet factor superfamily receptor; lymphocyte; chemotactic;
KW   Inflammation; inflammatory disease; arthritis; emphysema; cystic;
KW   fibrosis; colitis; bronchitis; meningitis; therapeutic; ss.
OS   Homo sapiens.
FH   Key               Location/Qualifiers
FT   cds               369..1487
FT                     /*tag= a
```

```
.  PN   WO9428931-A.
   PD   22-DEC-1994.
   PF   07-JUN-1994; U06380.
   PR   11-JUN-1993; US-076093.
   PA   (GETH ) GENENTECH INC.
   PI   Chuntharapai A, Hebert C, Kim KJ, Lee J;
   PT   Treatment of inflammatory disorders - by administering an
   PT   antibody capable of binding a platelet factor 4 superfamily
   PT   receptor polypeptide
   PS   Disclosure: Page 56-58; 83pp; English.
   CC   2 PF4AR members were identified by probing lambda libraries from
   CC   human monocyte-like cell line HL-60 and human peripheral blood
   CC   lymphocytes using a large fragment of IL-8 receptor DNA (full
   CC   sequence given in Q80520). The nucleotide sequences of the 2
   CC   PF4ARs are given in 080521 and 080522, and their respective
   CC   amino acid sequences in R68812 and R68813.
   SQ   Sequence 1679 BP; 327 A; 532 C; 457 G; 363 T;
```

Query Match 7.5%; Score 107.6; DB 1; Length 1679;
Best Local Similarity 47.0%;
Matches 409; Conservative 0; Mismatches 449; Indels 12; Gaps 2;
Pred. No. 2.8e-20;

```
Qy   277  TTCAAGTTCATCCTCGTCGTGTGACTATCCAGTGTCTTTGTTGTCGTTGGGCCTT 336
          |||||| |||||| ||| ||||  |||||||||  |||| ||||||||||||||
Db   513  TTCAAGGCCGTGTTCCGTGCCCGTGGCCTACAGCCTCATCTTCCTCCTGGGCCTTGGGCCTT 572

Qy   337  AACGCCCCAACCCTATGGGCTCTTCATCTTCCGCCTCCGACCCTGGGATGCAACGGCCACC 396
          ||||||||| |||||||||||||||||||||||||||| |||||| |||||| || ||
Db   573  AACGTCCTGTGTGCTGTGATCCCTGAGGCGACACGCAGTTCACGAGACC 632

Qy   397  TACAGTTCCACCTGGCATTGTCAGACACCTTGTATGTGCTGTCGCTGCCCACCCTCATC 456
          |||||||||||||||||| |||| ||||||||||| |||||||| || |||||||||
Db   457  TACATTATCAGCCCACACCACTGGCCTTTGGCCACTGGAGATCTGCAAGTTCGAAGATCTGCC 516

Qy   517  TTTCTTTTCTATTGGAAACCTACTACCACCTGCGTGTCCTTTTCCCTCACGTGGCATCAAGCGTGCAC 576
          ||||| ||||||| |||||||||||| ||||||||||| ||||||
Db   693  GCCGAGGGCTGTGTGGG----------CTGGGTCCTGGGGAGACTTCCTCTGCAAAACTGTGATT 746

Qy   577  CGCTACCTGGGCCATCTGCCACCACTTCGGGCGACTAGGCTGGGGCGCCCTTGCCCTTGCCA 636
          |||||||||||| |||| ||||||||||| |||| |||| ||||||||
Db   807  CGCTACCTGGGCCATTGTCCACGCCGTCCATGCCTACCGCCACCGCCGCCTTCCTCCCATC 866

Qy   637  GGCCTTCTCCTGCCTGGCCAGTTTGGTTGGTTGCGTAGCCGGCGTGGCGTGCTCGCCAGCCAACCTGTTC 696
```

```
Db   867  CACATCACCTGTGGGACCATCTGGGCTGGTTGCC...GAGAACACAGCAGAAAAAGGCATCCTGTTCACCTCGC... 926

Qy   697  TTTGTCACAACCAGCAACGAAAAGGGACACCACCGTCCTGTGCCATGACACCACTCGGCCTGAA 756
          ||||| |||| ||| ||| || |||| || || ||||| |||| || ||
Db   697  TTCGCCAAAGTCAGCCAAGGCCATCACAACAACTCCCCTGCCCACGGTTGCACCTTCATCCCAA 756

Qy   757  GAGTTTGACCACTATGTGCACTTCAGCTCGGGGGGGTCATGGGGCTGCTCTTTGGCGTGCC 816
          ||| ||||||||||||| |||||||||||||| ||| ||||| ||| |||| ||| ||
Db   987  GAGAACCAAGCAGAGAAAGCCATGCCTGGTTCACCTCCCGATTCCTCTACCATGTGGGCAGG 1046

Qy   817  TGGCCTGGTGCACTTGTTGTTGCTATGAGGACTCATAGGCTCGTCGCCTGTATCAGCCCTTGCCA 876
          |||||| |||||| |||||||| || || ||||||| ||||| |||| ||||||
Db   817  TTCCTGCTTGGCCCCCATGCTGGTGAATGGGGCTTACGTGGGGGGTCACAAGGTTGGC 876

Qy   877  GGCTCTCCACAGTCGTCTTCTCCGCTCTCTCCGCACCATAGCTGTTGCTGCTGACT 936
          |||| ||| ||||||||||||||| ||||| || |||||||| ||| || ||
Db   1107  CAGGCCCACGGCGCCCTCAGCCGGCCAGAAGCAGTCAGGGTGGCCATCCTGGTGACAAGC 1166

Qy   937  GTCTTTGCTGTCTGCTTCGGTCCTTTCCACATCACCGCGACCATTACTACCGGGCAGG 996
          |||||| || ||||||||||||||||||||||| ||||||| ||||||||||| ||
Db   1167  ATCTTCTTTCCCTGTCGGTCACCCTACCACACCGTCATCTTCCTGGACGCCCTGGGCGAGG 1226

Qy   997  CTGTTGGAAGCTGAC----------TGCCGAGTACTAGAACATTGTCAACGTGGTCTATAAAGTG 1050
          |||||||||||||||          || ||||| || ||||| || |||| |||||
Db   1227  CTGAAGGCCGTGGACAATACCTGCAAGCTGAATGGGGCTCTCTCCCCGGTGGCCATCACCATG 1286

Qy   1051  ACTCGGCCCCCTGCCCAGTGCCAACAGCTGCTGGAGTCCTGGATCCTGCTTCTTGCTGCTTCACTCTTGCTCACTGGG 1110
          |||||||||||||| ||||||||| ||| || |||| || || ||||| |||||
Db   1287  TGTGAGTTCCTGGGCCTGGCCCTGGCCCCCATGCTGCTCCAACCCATGCTCTACACACTTTCGCCGGC 1346

Qy   1111  GACAAATATCGACGTCAGCTCCGTCAGCTC 1140
          ||||||||| ||||||| |||||||||||||
Db   1347  GTGAAGTTCCCAGTGACCTGTCGCGGCTC 1376
```

RESULT 12

```
Q99009
ID   Q99009 standard; cDNA; 1679 BP.
AC   Q99009;
DT   26-MAR-1996 (first entry)
DE   Chemokine superfamily receptor coding sequence.
KW   Interleukin; IL-8; inflammation; psoriasis; dermatitis;
KW   rheumatoid arthritis; inflammatory bowel disease;
KW   chronic lung inflammation; treatment; antibody;
KW   affinity purification; detection; ss.
OS   Homo sapiens.
PN   US5440021-A.
PD   08-AUG-1995.
PF   29-MAR-1991; 677211.
PR   29-MAR-1991; US-677211.
PR   25-FEB-1994; US-202056.
PA   (CHUN/) CHUNTHARAPAI A.
PA   (HEBE/) HEBERT C.
PA   (KIMK/) KIM K J.
PA   (LEEJ/) LEE J.
PI   Chuntharapai A, Hebert C, Kim KJ, Lee J;
DR   WPI; 95-283151/37.
DR   P-PSDB; R92239.
PT   New antibodies against interleukin 8 type B receptor - used to treat
PT   or prevent inflammation, also for detecting receptor expression and
PT   purifien.
PS   Example 2; Columns 51-54; 62pp; English.
CC   Antibodies directed against the interleukin-8 receptor B can be used
CC   to treat or prevent inflammation e.g. psoriasis, dermatitis,
CC   rheumatoid arthritis and particularly inflammatory bowel disease and
CC   chronic lung inflammation. When immobilised, these antibodies may
CC   be used to detect interleukin-8 receptor B expression in cells and
CC   tissues and for affinity purification of interleukin-8 receptor B
CC   from cells. This sequence is an additonal chemokine superfamily
CC   receptor which was identified by probing lamda libraries of genomic
CC   DNA from a human monocyte-like cell line (L-60) and human peripheral
CC   blood lymphocytes using a large fragment of the interleukin-8 type
```

CC   A receptor DNA (See Q99006).
SQ   Sequence 1679 BP; 327 A; 532 C; 457 G; 363 T;

Query Match   Best Local Similarity   47.0%;
Matches 409;   Conservative   7.5%;   Score 107.6; DB 1; Length 1679;
               Pred. No. 2.8e-20;   Mismatches 449; Indels 12; Gaps 2;

```
Qy    277 TTCAAGTTCATCCTGCCTGCCTGTGAGCTATGCAGTTGTCTTTGTGCTGGGGTTGGGCTT 336
          ||                    |   ||  ||     |   | |    |||   ||  ||
Db    513 TTCAAGGCGTGTGTCGTGCCCGTGCCTACCAGCCTCATCTTCCTCTCGGGGCGTGATCGGC 572

Qy    337 AACGCCCCAAACCCTATGGGCTCTTCATCTTCCGCCTTCCGACCTGGGATGCAAACGGCCACC 396
          ||||| ||            |||     |      || |||     |||||||     |||
Db    573 AACGTCCTGGTGCTGGTGGTGATCCTGGAGGGGCACGGCGAGACACGCAGTTCACGGAGACC 632

Qy    397 TACAGTTCCACCTGGCATTGTCAGACACCTTTGTATGTGTCGCTGCCCACCTCATC 456
          |  |||||||||||||||| |||     |||||||||      |||||||   ||
Db    633 TTCCGTGTTCCACCTGGGCCGGGGGACCCTCGTTGGTCTTCATCTTGCCCTTTGCCCTGTG 692

Qy    457 TACTATTATGCAGCCACCAACACCACTGGGCCCTTTGGCACTGAGGATCTGCAAGTTCGTCCGC 516
          |||      |                  |||  |||      ||||||||||| |||||
Db    693 GCCGAGGGGCTCTGTGGG------CTGGGTCCTGGGGACCTTCCTTCTGCAAAACTGTGATT 746

Qy    517 TTCTTTTTCTATTCGGAACTTCTACTGACAGTGTCTTTTTTCCTCACCTGCATCAGGGTGCAC 576
          ||| ||        |||     |||||| |||||||||||||   |||||||    |||||
Db    747 GCCCTGCACAAAGTCAACTTCTACTGCAGCAGCTCCGTCCTCGGGCCTGCATCGCGCTGGAC 806

Qy    577 CGCTACCTGGGCATCTGCCACCCACTTCGGGCACAGCTGGGGGGGCGCCCTCGGCTCGGCA 636
          ||||||||||||||||   |||||||||||||     | ||| ||||||  ||||| ||
Db    807 CGTTACCTGGCCATTGTCCACGCGCTACCGACCCGCCCTCCTCCATC 866

Qy    637 GGGCTTCTCGCCCTGGCAGTTTGGTTGGTTGGTAGCCGGGGTGCCCAACCTGTTC 696
          ||       |   |      |||   ||||| ||   |||||||||||
Db    867 CACCACCTGTGGGACCATCTGGTGGTCCTTGGCTTCCTTGGCCAGAGATTCTC 926

Qy    697 TTTGTCACAACCACCAACAAAGGGACCACCGTTCTGTGCCATGACACCACTCGCCTGAA 756
          || ||||     ||| ||  |||||||||||||||  |||||||||||||
Db    927 TTCGCCAAAGTCAGCCAAGCCATCACAACAACTCCCGCACGTTGCCACCTTCTCCCAA 986

Qy    757 GAGTTTGACCACTATGTGCACTTCAGCTCGGCGGGTCATGGGGCGTCATGGGGCGTGCCC 816
          ||||||||        |||||||||||||||||  ||||||| |||  | ||||||
Db    987 GAGAACCAAGCAGAAACGCATGCCTGGGTTCACCTCCGATTCTCTACCACCATGTGGGGGA 1046

Qy    817 TGCCCTGGTGCACTCTGTTGTTTGCTATGGACTCATGGCCTCGTGCCCGTATCAGCCCA 876
          ||                            |||||||||||    |||||  |||
Db   1047 TTCCTGCTTGCCCAATGGCTGGTGATGGGGTCGTACGTGGGGGTAGTGCACAGGTTGCCC 1106

Qy    877 GGGCTCTGCACAGTCGTCTTCTCGCCTCCGCTCCGCCACCAATAGCTGTGGTGCTGACT 936
          || |||||||||||    | |       |   |  ||||||||||||||
Db   1107 CAGGCCCACCGGGCCCCTCAGCGGCGGAAAGCCAGTGGGCCATCCTGGTGCACAAC 1166

Qy    937 GTCTTTGCTCGTCTGCTCTTGTGTCCATGACCACCACTCGCCTGAA 996
          ||  |   | |  |           ||||     |   ||||
Db   1167 ATCTTCTTCTCCTGCTGCTTTCCACTCACCACCGGACCACCCGGGCGAGG 1226

Qy    997 CTGTTGGAAGCTGAC-----TGCCGAGTACTGAACAATTGTCAAGGTGGTTCTATAAAGTG 1050
          |||  |                |   |||||| | ||||| ||
Db   1227 CTGAAGGGACCGTGGACAATACCTGGTTCACCTCTCCCGGTGGCCATCACCATG 1286

Qy   1051 ACTCGGCCCCCTGGCCACGTCGCCAACAGCGCCTGCCTTGGGATCCTGCTTCACTGG 1110
          |||||||||||||     |   ||||||||  | ||      ||||| |
Db   1287 TGTGAGTTCCTGGGCCTGGGCCCAGTGCTCCTCAACCCCTACACACTTTCGCCGGC 1346

Qy   1111 GACAAAATATCGAGGTCAGCTCCGTCCGGCTCAGCTC 1140
          | |        ||||||||||||| ||| |||||  |
Db   1347 GTGAAGTTCCGCAGTGACCTGTCGCGGGTC 1376
```

PT   New G protein-coupled receptor - useful for diagnosis, treatment and
PT   prevention of neutropaenia, agranulocytosis, infection and cancer
PS   Claim 11; Fig 1; 4pp; English.
CC   This cDNA clone encodes a novel human G protein coupled receptor,
CC   termed P2Y11 (see W97842), that has selective affinity for ATP. A
CC   human cDNA placenta cDNA library was screened with a human P2Y4
CC   probe. Of 9 clones obtained, 3 corresponding to a partial sequence
CC   encoding a new G protein coupled receptor displaying about 30%
CC   identity with other P2Y receptors. This partial sequence was used
CC   as a probe to screen a human genomic DNA library. 4 Overlapping
CC   genomes clones were isolated. Mapping and sequencing showed the
CC   new gene contained an intron at the 5' end of the coding region.
CC   The 4 clones contained the entire open reading frame for the new
CC   receptor, designated P2Y11. The invention also provides vectors,
CC   transformed cells, anti-P2Y11 antibodies, nucleic acid probes,
CC   pharmaceutical compositions comprising such products and transgenic
CC   animals. Antisense nucleotides (claimed) that hybridise to mRNA
CC   are used to decrease activity of P2Y11, while specific antibodies
CC   are used to block binding of P2Y11 to its ligand. Probes are used
CC   in hybridisation assays to detect expression of P2Y11 at the RNA
CC   level, while antibodies are used similarly at the protein level in
CC   standard immunoassays, particularly for diagnosis of leukaemia.
CC   The transgenic animals are used to determine the effects of varying
CC   levels of P2Y11 expression. These animals, and host cells, are
CC   used in drug screening methods to identify (antagonists) that are
CC   potentially useful for treatment or prevention of disorders
CC   associated with excessive or inadequate receptor activity,
CC   specifically neutropaenia, agranulocytosis, infections and cancer.
CC   Host cells are also used to produce recombinant P2Y11.
SQ   Sequence 2427 BP; 464 A; 790 C; 774 G; 399 T;

Query Match   Best Local Similarity   55.9%;
Matches 219;   Conservative   7.3%;   Score 104; DB 1; Length 2427;
               Pred. No. 3.2e-19;   Mismatches 170; Indels 3; Gaps 1;

```
Qy    310 GTTGTCTTTGTGCTGGGGTTGGGCCTTAAGCGACCCAACCCTATGGGCTCTTCATCTTCCG- 369
          ||||                 | |           |||||||||||||||||||||||||
Db   1408 GTTGAGTTCCTGGTGGGCGGTGGCCGTGGCCAGCAAGGCCGTGGCCTGTACCGGCTTCAGACACC 1467

Qy    369 --CCTCCGACCCTGGGATGCAAACGGCCACCTACATGTTCCACCTGGCATTGTCAGACACC 426
            ||  | || |  ||||||| |||||   ||||    ||||||     ||||
Db   1468 AAGGACGCGCCCATGCACCCCGCGTGGTCTTCTTCTCGTGCCAGTGGCGACTCAGGGACCTG 1527

Qy    427 TTGTATGTGTCGCTGCCCACCCTCATCACTATTATGCAGCCCACCAACCACTGGGCCGCC 486
          ||    | || | |    ||||||||||||||||||||||||    ||| |||
Db   1528 CTCTCGCGCTCGACGGCTGCCCCGGGCGCCCTGGAGACCCTCTCTTCACCCTACCACTGGGCAGC 1587

Qy    487 TTTTGGCACTGAGATCTGAAAGTTCGTCCGGCTTTCTTTCATTTCTTATTGGAACTTC 546
          ||   ||     |         ||||||||||||||||||||||||||||||||||
Db   1588 TATGGGAGGAGGCGCGCCGGGCCTGGAAGCCGCATCGTGCACCGTCCACCGTGGGCAAC 1647

Qy    547 GTCCTTTTCTCACCTGCATCAGGGTGCACCGCTACCTGGGCATCGCCACCCACTTCGG 606
          ||| |     |||||| ||  |||||||||||||||| |||||||||||||||||
Db   1648 GTCATCTTCATCACCTCAACCGGTACCTACCGTGGCCACGTCGCACCCTTCTTC 1707

Qy    607 GCACTACGCTGGGGGCGCCCTCGGCCGAGGCCTTCTGCCTGGCCAGTTTGGTTGGTC 666
```

ID   X07369 standard; cDNA; 2427 BP.
AC   X07369;

Db 1708 GCCCGAACCACCTGCGACCCAAGCACGCCGCTGGGCCGTGAGCGCCTGCCTCCGGCTGGGGTCCTG 1767

Qy 667 GTAGCCGGCTGCCTCGCTGCGCCCAACCTGTTCTT 698
       ||| |||| | |||||| |||||| || |||
Db 1768 GCCGCCCTGCTGGCCATGCCACACTCAGCTT 1799

RESULT 14
Q84557
ID Q84557 standard; DNA; 1477 BP.
AC Q84557;
DT 22-AUG-1995 (first entry)
DE Murine C140 receptor genomic DNA.
KW G-protein-coupled receptor; G-protein; C140 receptor; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 232..1188
FT /*tag= a

PN WO9503318-A.
PD 02-FEB-1995.
PF 26-JUL-1994; U08536.
PR 26-JUL-1993; US-097938.
PA (CORT-) COR THERAPEUTICS.
PI Scarborough RM, Sundelin J;
WPI 95-075182/10.
DR P-PSDB; R66920.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Disclosure; Fig 1; 57pp; English.
CC A mouse cosmid genomic library (obtd. from Dr R.A Wetsel, Washington
CC Univ. School of Medicine, St Louis, Missouri) was screened with two
CC 32P-labeled oligos corresp. to bp 190-249 and 742-801 of the bovine.
CC A substance K receptor cDNA. In one of the clones isolated (C140) the
CC hybridising region was localised to a 3.7 kb PstI fragment. This
CC fragment was subcloned into pBluescript vector. The hybridising and
CC adjacent regions were sequenced. The nt sequence and the deduced AA
CC sequence are given in Q84557 & R69920 respectively. The published
CC DNA sequence appears to have been printed with the last two lines
CC in the wrong order. This has been corrected in this Geneseq entry.
SQ Sequence 1477 BP; 341 A; 414 C; 320 G; 402 T;

Query Match 6.9%; Score 98.6; DB 1; Length 1477;
Best Local Similarity 57.2%;
Matches 179; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 287 TCCTGCTGCCTGTGAGCTATGCAGTTGNCTTTGTGCTCGGGCCTTAACGCCCCAA 346
Db 452 TCTTTCTTCCGGTCGTCTACATATTATTGTGTTTGATNGTTTGTCCCAGTAATGGCATGG 511

Qy 347 CCCTATGGCTCTTCATCTTCCGGCCTCCGACCCGGGATGCAACGGCCACCTACATGTTCC 406
Db 512 CCCTCTGGATTCCTTTCCGGATGCAAGGAAACACCCCGCGCCGTGATTTACATGGGCA 571

Qy 407 ACCTGGCARTGTCAGACACCTTGRATGNTGCTGTGTGGCTGCCACCCTCANCTCATTATG 466
Db 572 ACCTGGCCTTGGGCGACCTCCTCTTGTCATCTGGTTCCCCTGAAGARTCTCCTTACCACC 631

Qy 467 CAGCCCACACAACCACTGGCCCTTTGGCACTGAGARTCTCGAAGTTGTCTCCCGCTTTCTTTTTCT 526
Db 632 TACAATGGCAACAACTGGGTCTACGGGAGGCCCTGTGCAAGGTGCTCANTGGCTTTTTTCT 691

Qy 527 ATTGGAACCTCRACTGCAGTGTCCTTTTCCTCACCTGCRATCACGTGCACCGTACCTGG 586
Db 692 ATGGTAACATGTRATTGCTCCATCCTCTTCATGACCTGCCTCAGCGTGCAGAGGTACTGG 751

Qy 587 GCATCTGCCACCC 599
       |||
Db 752 TGARCGTGAAACC 764

RESULT 15
Q84559
ID Q84559 standard; cDNA; 2732 BP.
AC Q84559;
DT 22-AUG-1995 (first entry)
DE Murine C140 receptor cDNA.
KW G-protein-coupled receptor; G-protein; C140 receptor; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 73..1272
FT /*tag= a

PN WO9503318-A.
PD 02-FEB-1995.
PF 26-JUL-1994; U08536.
PR 26-JUL-1993; US-097938.
PA (CORT-) COR THERAPEUTICS.
PI Scarborough RM, Sundelin J;
WPI 95-075182/10.
DR P-PSDB; R66922.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Claim 1; Fig 10; 57pp; English.
CC A cDNA library from a mouse stomach was constructed in lambda gt10
CC and screened with a probe encompassing the C140 genomic clone (see
CC Q84557). A single phage clone was isolated and cut with EcoRI. The
CC insert was cloned in pBluescript and pSG5 and sequenced. The
CC complete nt sequence and deduced AA sequence is given in Q84559 &
CC R66922. 5' RACE resulted in the addition of only 27 bps to the 5'
CC end. The 5' end of the apparent coding region differs from the 5'
CC end of the ORF of genomic DNA; it is believed that the 5' end of
CC the cDNA sequence is correct.
SQ Sequence 2732 BP; 650 A; 676 C; 633 G; 773 T;

Query Match 6.9%; Score 98.6; DB 1; Length 2732;
Best Local Similarity 57.2%;
Matches 179; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 287 TCCTGCTGCCTGTGAGCTATGCAGTTGNCTTTGTGCTCGGGCCTTAACGCCCCAA 346
Db 305 TCTTTCTTCCGGTCGTCTACATATTATTGTGTTTGATNGTTTGTCCCAGTAATGGCATGG 364

Qy 347 CCCTATGGCTCTTCATCTTCCGGCCTCCGACCCGGGATGCAACGGCCACCTACATGTTCC 406
Db 365 CCCTCTGGATTCCTTTCCGGAACGAAGGAAACACCCCGCGCCGTGATTTACATGGGCCA 424

Qy 407 ACCTGGCARTGTCAGACACCTTGRATGNTGCTGTGTGGCTGCCACCCTCANCTCATTATG 466
Db 425 ACCTGGCCTTGGGCGACCTCCTCTTGTCATCTGGTTCCCCTGAAGARTCTCCTTACCACC 484

Qy 467 CAGCCCACACAACCACTGGCCCTTTGGCACTGAGARTCTCGAAGTTGTCTCCCGCTTTCTTTTTCT 526
Db 485 TACAATGGCAACAACTGGGTCTACGGGAGGCCCTGTGCAAGGTGCTCANTGGCTTTTTTCT 544

Qy 527 ATTGGAACCTCRACTGCAGTGTCCTTTTCCTCACCTGCRATCACGTGCACCGTACCTGG 586
Db 545 ATGGTAACATGTRATTGCTCCATCCTCTTCATGACCTGCCTCAGCGTGCAGAGGTACTGG 604

Qy 587 GCATCTGCCACCC 599
       |||
Db 605 TGARCGTGAAACC 617

Search completed: December 29, 1999, 07:20:25
Job time: 8929 sec

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 04:48:33 ; Search time 601.47 Seconds
(without alignments)
10158.894 Million cell updates/sec

Title: US-09-077-173A-1
Perfect score: 1429
Sequence: 1 AAGGGAGCTTGGGTAGGGGC..........GCCACGGGAGGCAGGGATATC 1429
Scoring table: IDENTITY_NUC
Searched: 780561 seqs, 2137953050 residues
Database: GenEmbl:*
Word size: 0

Number of hits that pass the threshold : 1561122

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pri:*
10: gb_pr1:*
11: gb_pr2:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_htg:*
19: em_hum1:*
20: em_hum2:*
21: em_in1:*
22: em_om:*
23: em_or:*
24: em_ov:*
25: em_pat:*
26: em_ph:*
27: em_pl:*
28: em_ro:*
29: em_sts:*
30: em_sy:*
31: em_un:*
32: em_vi:*
33: gb_htg1:*
34: gb_htg2:*
35: gb_in1:*
36: gb_in2:*
37: em_ba1:*
38: em_ba2:*
39: em_hum3:*
40: em_hum4:*
41: gb_pr4:*
42: gb_htg3:*
43: gb_htg4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1424.2	99.7	1429	9 5	A62985	A62985 Sequence 1
2	1424.2	99.7	1429	9	HSP2Y4	X91852 H.sapiens P
3	1363.6	95.4	1651	11	HSU40223	U40223 Human uridi
4	1228.8	86.0	1232	9	HSGPCRE	X96597 H.sapiens g
5	802.2	56.1	2684	12	RNP2Y4RG	Y14705 Rattus norv
6	794.4	55.6	1977	12	RNP2Y4REC	Y11433 R.norvegicu
7	416.4	29.1	1977	4	AF031897	AF031897 Meleagris
8	379.2	26.5	2025	11	HSU07225	U07225 Human P2U n
9	379.2	26.5	2025	11	HP2U=P2U nu	S74902 Human P2U n
10	359.2	25.1	2011	12	RATP2U	U56839 Rattus norv
11	357.6	25.0	1688	12	RNU56839	U56839 Rattus norv
12	352	24.6	1842	5	RNU09402	U09402 Rattus norv
13	352	24.6	1842	5	I34573	I34573 Sequence 1
14	352	24.6	1842	5	I36524	I36524 Sequence 1
15	346.6	24.3	1131	12	I76280	I76280 Sequence 1
16	343.4	24.0	2138	12	S83099	S83099 P2U recepto
17	326.8	22.9	2361	4	MUSP2UREC	L14751 Mouse ATP r
18	218.6	15.3	984	5	XLP2Y8	X99953 X.laevis mR
19	218.6	15.3	1571	9	E12487	E12487 Human cDNA
20	218.6	15.3	1507	10	HSP2Y6	X97058 H.sapiens m
21	218.6	15.3	1539	10	HSU52464	U52464 Human P2 pu
22	218.6	15.3	1832	11	AF007891	AF007891 Homo sapi
23	218.6	15.3	984	5	AF007892	AF007892 Homo sapi
24	210.8	14.8	2352	11	E12485	E12485 Mouse cDNA
25	209.2	14.6	1922	12	AF007893	AF007893 Homo sapi
26	208.2	14.6	1163	4	RATNGPCP2R	D63665 Rat mRNA fo
27	195.8	13.7	1498	4	GDP2Y3	X98283 G.domesticu
28	195.8	13.7	1074	4	GDATPREC	X73268 G. domestic
29	187.8	13.1	1192	4	AF065555	AF065555 Meleagris
30	187.8	13.1	1180	4	AF012103	AF012103 Meleagris
31	185.8	13.0	1180	4	MGU09842	U09842 Meleagris g
32	185.8	13.0	1666	4	BTP2YRECP	X87628 B.taurus mR
33	167.4	11.7	1308	3	BTU34041	U34041 Bos taurus
34	167.4	11.7	16178	3	HSA569	AJ006945 Human P2Y
35	147.4	10.3	1426	10	HSU42029	U42029 Human P2Y1
36	147.4	10.3	2424	10	HSU42030	U42030 Human P2Y
37	147.4	10.3	1312	11	S81950	S81950 P2 purinoce
38	141.2	9.9	3204	12	RNU22830	U22830 Rattus norv
39	138	9.7	400	3	AF005153	AF005153 Bos tauru
40	136.2	9.5	1293	12	MM22829	MM22829 Mus musculu
41	134.6	9.4	3055	9	HSATPRMR	Z49205 H.sapiens m
42	134.6	9.4	12630	12	MMU245636	AJ245636 Mus muscu
43	134.6	9.4	3055	13	G28604	G28604 human STS S
44	128.4	9.0	2070	9	Y12546	Y12546 H.sapiens m
45	128.4	9.0	2889445	42	AC010878	AC010878 Homo sapi

ALIGNMENTS

RESULT 1
A62985

```
LOCUS       A62985      1429 bp    DNA             PAT       12-MAR-1998
DEFINITION  Sequence 1 from Patent WO9719170.
ACCESSION   A62985
NID         g3716857
VERSION     A62985.1  GI:3716857
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1429)
  AUTHORS   Communi,D., Pirotton,S., Parmentier.M. and Boeynaems,J.
  TITLE     RECEPTOR AND NUCLEIC ACID MOLECULE ENCODING SAID RECEPTOR
  JOURNAL   Patent: WO 9719170-A 1 29-MAY-1997;
            EUROSCREEN S A (BE)
FEATURES             Location/Qualifiers
     source          1..1429
                     /organism="unidentified"
```

```
CDS             181..1278
                /db_xref="taxon:32644"
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAA03704.1"
                /db_xref="PID:e1265092"
                /db_xref="PID:g3716858"
                /db_xref="GI:3716858"
                /translation="MASTESSLLRSLGLSPGPGSSEVELDCWFDEDEKFILLPVSYAV
                VFVLGLGLNAPTLWLFIFRLRPWDANATYLMHLALSDTLYVLSLPLLIYYAAHHWP
                FGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVW
                LVVAGCLVPNLFFVTTSNKGTTVLCHDTTRPEEFDHYVHFSSAVMGLLFGVPCLLTLV
                CYGLMARRLYQPLPGSAQSSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYILARLLEA
                DCRVLINVVNVYKVTRPLASANSCLDPVLYILLTGDKYRROLRQLCGGKPQPRTASS
                LALVSLPEDSSCRWAAIPQDSSCSTPRADRL"
```

```
BASE COUNT    242 a    451 c    379 g    357 t
ORIGIN
```

```
Query Match        99.78;  Score 1424.2;  DB 5;  Length 1429;
Best Local Similarity 99.8%;  Pred. No. 0;  Mismatches 3;  Indels 0;  Gaps 0;
Matches 1426;  Conservative 0;  Mismatches 3;  Indels 0;  Gaps 0;
```

```
Qy      1 AAGGGAGCTTGGGTAGGGGCCAGGCTAGCCTAGTGCTGCACCCAGATGGCCTTCGTCAGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGGGAGCTTGGGTAGGGGCCAGGCTAGCCTAGTGCTGCACCCAGATGGCCTTCGTCAGCT 60

Qy     61 CTCCCTAGTGCTTCAACCACTGCTCTCCCTGCTCTACTTTTTTTTTGCTCCAGCTCAGGGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTCCCTAGTGCTTCAACCACTGCTCTCCCTGCTCTACTTTTTTTTTGCTCCAGCTCAGGGAT 120

Qy    121 GGGGGGGGGGCAGGGAAATCCTGCCACCCTCACTTCCCCTTCCCAATCTCCAGGGGGGCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGGGGGGGGCAGGGAAATCCTGCCACCCTCACTTCCCCTTCCCAATCTCCAGGGGGGCC 180

Qy    181 ATGGCCAGTACAGAGAGTCCTCCCTGTTGAGATCCCTAGGCCTTCCAGCGTCCTGGCAGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGGCCAGTACAGAGAGTCCTCCCTGTTGAGATCCCTAGGCCTTCCAGCGTCCTGGCAGC 240

Qy    241 AGTGAGGTGGAGCTGAGCTGAGAGCCGAGGAATTCAAGTTCATCCGCTGCCCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTGAGGTGGAGCTGAGCTGAGAGCCGAGGAATTCAAGTTCATCCGCTGCCCTG 300

Qy    301 AGCTATGCAGTTGTCTTTGTGCTGGGCCTTAACGCCCCAACCCTATGGCTCTTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCTATGCAGTTGTCTTTGTGCTGGGCCTTAACGCCCCAACCCTATGGCTCTTC 360

Qy    361 ATCTTCCGCCTTCCGACCCTGGGATGCAAACGGCCACCTACACATGTTCACCTGTCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATCTTCCGCCTTCCGACCCTGGGATGCAAACGGCCACCTACACATGTTCACCTGTCA 420

Qy    421 GACACCTTGTATGTGCTGGCTGCCTGCCACCCTCACTCCTACTATTATGCAGCCCACAACCAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GACACCTTGTATGTGCTGGCTGCCTGCCACCCTCACTCCTACTATTATGCAGCCCACAACCAC 480

Qy    481 TGGCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTCTTTTTCTATTGGAACCTCTAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGGCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTCTTTTTCTATTGGAACCTCTAC 540

Qy    541 TGCAGTGTCCTTTTCCTCACCTGCATCAGCGTGCACCGCTACCTGGGCATCTGCCACCCA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGCAGTGTCCTTTTCCTCACCTGCATCAGCGTGCACCGCTACCTGGGCATCTGCCACCCA 600

Qy    601 CTTCGGGCACTACGCTGGGGCCGCCGCCTTCTCGTCCTGGGCCTGCCAGTTTGG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTTCGGGCACTACGCTGGGGCCGCCGCCTTCTCGTCCTGGGCCTGCCAGTTTGG 660

Qy    661 TTGGTCGTAGCCGGCTGCCCTCGTGCCCACCCTGCACACCGGAACAACAAAGGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTGGTCGTAGCCGGCTGCCCTCGTGCCCACCCTGCACACCGGAACAACAAAGGG 720

Qy    721 ACCACCGTCCTGTGCATGACACCTGGCCTGAGAGAGTTTGACCACTATGTGCACTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACCACCGTCCTGTGCATGACACCTGGCCTGAGAGAGTTTGACCACTATGTGCACTTC 780
```

```
Db    721 ACCACCGTCCTGTGCATGACACCTGGCCTGAGAGAGTTTGACCACTATGTGCACTTC 780
Qy    781 AGCTCCGCGGTCATGGGGCTGCTCTTTGGCGTGCCCTGTCTGCTCACTCTGTTGTGTTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    781 AGCTCCGCGGTCATGGGGCTGCTCTTTGGCGTGCCCTGTCTGCTCACTCTGTTGTGTTAT 840

Qy    841 GGACTCATGGCCCGTCGCCTGTATCAGCCCCTTGCCAGCGCTCGCACAGTCGTCTTCTCGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGACTCATGGCCCGTCGCCTGTATCAGCCCCTTGCCAGCGCTCGCACAGTCGTCTTCTCGC 900

Qy    901 CTCCGCCTCTCCGCCACCATAGCTGTTGGTCCTGACCTTCGGTGCTGTCTTTGCTGTTGTGCT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CTCCGCCTCTCCGCCACCATAGCTGTTGGTCCTGACCTTCGGTGCTGTCTTTGCTGTTGTGCT 960

Qy    961 TTCCACAATCACCCGCACCATTTACTTACCTGTGGTGGTGCTGACTGTTGGAAGCTGAGTA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCCACAATCACCCGCACCATTTACTTACCTGTGGTGGTGCTGACTGTTGGAAGCTGAGTA 1020

Qy   1021 CTGAACAATTGTCAACGTGGTCTATAAAGTGACTCGGCCCCTGCCCAGTGCCAACAGCTGC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTGAACAATTGTCAACGTGGTCTATAAAGTGACTCGGCCCCTGCCCAGTGCCAACAGCTGC 1080

Qy   1081 CTGGATCCTGTGCTCTACTTGCTCACTGGGGACAAAATATCGACGTCAGCTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTGGATCCTGTGCTCTACTTGCTCACTGGGGACAAAATATCGACGTCAGCTC 1140

Qy   1141 TGTGGTGGTTGGGTGGCAAGCCCCAGCCCCACGCGGCGGTGCCGCTGGCACTAGTGTCCCTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGTGGTGGTTGGGTGGCAAGCCCCAGCCCCACGCGGCGGTGCCGCTGGCACTAGTGTCCCTG 1200

Qy   1201 AGGGCAGATAGGATTGTAACACCGGGAAGTGAGAGAAAAGGGGATGAGTGAGG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AGGGCAGATAGGATTGTAACACCGGGAAGTGAGAGAAAAGGGGATGAGTGAGG 1260

Qy   1261 GCAGAGGTGAGGGAACCCAATAGTGAATACCTGGTAAGGTGGCTTCTCCCTCTTTTCCAGGC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GCAGAGGTGAGGGAACCCAATAGTGAATACCTGGTAAGGTGGCTTCTCCCTCTTTTCCAGGC 1320

Qy   1321 TCTGGAGAGAAGCCCTCACCCTGAGGGGTTGCCACGGAGGCAGGGATATC 1429
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCTGGAGAGAAGCCCTCACCCTGAGGGGTTGCCACGGAGGCAGGGATATC 1429
```

```
RESULT 2

HSP2Y4

LOCUS       HSP2Y4              1429 bp    DNA             PRI       10-APR-1996
DEFINITION  H.sapiens P2Y4 gene.
ACCESSION   X91852
NID         g1124904
VERSION     X91852.1  GI:1124904
KEYWORDS    P2Y4 gene; uridine nucleotide receptor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1429)
  AUTHORS   Communi,D., Pirotton,S., Parmentier,M. and Boeynaems,J.M.
  TITLE     Cloning and functional expression of a human uridine nucleotide
            receptor
  JOURNAL   J. Biol. Chem. 270 (52), 30849-30852 (1995)
  MEDLINE   96125055
REFERENCE   2  (bases 1 to 1429)
  AUTHORS   Communi,D.B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-1995) D.B.C. Communi, Institute of
            Interdisciplinary Research, U.L.B., Building C (local C5-145),
            Campus Erasme,Route de Lennik 808, B-1070 Brussels, BELGIUM
FEATURES             Location/Qualifiers
     source          1..1429
```

```
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="lambda Charon IVa"
gene            181..1278
                    /gene="P2Y4"
CDS             181..1278
                    /gene="P2Y4"
                    /codon_start=1
                    /product="uridine nucleotide receptor"
                    /protein_id="CAA62963.1"
                    /db_xref="PID:e204061"
                    /db_xref="PID:g1124905"
                    /db_xref="GI:1124905"
                    /db_xref="SWISS-PROT:P51582"
                    /translation="MASTESSLLRSIGLSGPGSSEVELDCWFEDEDFKFILLPVSYAV
                    VFVLGLINAPTIWLFIFRLRPWDATATVMFHLALSDIIYVLSLPTLIYYAAHNHWP
                    FGEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVW
                    LVVAGCLVPNLFFVTTSNKGTTVLCHDTTPEEEDHYVHRSSAVWGLLPGVPCLVTLV
                    CYLMARRLYQPLPGSASQSSSRSRSLRISIRRSIYIAVVLIVFAVCFVPHITRTIYYLARLLEA
                    DCRVLNIVNVVYKVTRPLASANSCLDPVLYLLTGSDKYRRQLRQLCGGGKPQPRIAASS
                    LALVSLPEDSSCRWAATPQDSSCSTPRADRL"
BASE COUNT    242 a    451 c    379 g    357 t
ORIGIN
```

```
Query Match      99.7%;      Score 1424.2;      DB 9;      Length 1429;
Best Local Similarity      99.8%;      Pred. No. 0;      Mismatches   3;      Indels   0;      Gaps   0;
Matches 1426;    Conservative    0;    Mismatches    3;    Indels    0;    Gaps    0;

Qy      1 AAGGGAGCTTGGGTAGGGGGCCAGGGCTTAGCGTTGAGTGTGCACCCAGATAGCGCTTCTTGTCTCTTGTCAGCTCAGCTT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGGGAGCTTGGGTAGGGGGCCAGGGCTTAGCGTTGAGTGTGCACCCAGATAGCGCTTCTTGTCTCTTGTCAGCTCAGCTT 60

Qy     61 CTCCCTAGTTGCTTCAACCACTGCTCTCCCCTGCCTCCCTGCCTCTACTTTTTTTTTTTGGCTCAGCTCAGGGAT 120
          ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     61 CTCCCTAGTTGCTTCAACCACTGCTCTCCCCTGCCTCCCTGCCTCTACTTTTTTTTTTTGGCTCAGCTCAGGGAT 120

Qy    121 GGGGGTGGGCCAGGGAGAAATCCTGCACCCTCACTTCTCTCCCCTTCCCATCTCCAGCTCCAGGGGGCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGGGTGGGGCCAGGGAGAAATCCTGCACCCTCACTTCTCTCCCCTTCCCATCTCCAGCTCCAGGGGGC 180

Qy    181 ATGGCCAGTACAGAAGTCCTCCCCTGTTGAGAATCCCTAGGCCTCAGCCCCAGGTCCTGGCAGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGGCCAGTACAGAAGTCCTCCCCTGTTGAGAATCCCTAGGCCTCAGCCCCAGGTCCTGGCAGC 240

Qy    241 AGTGAGGTGGGAGCTGGACTGTTGGTTTGATGAGGATTTCAAGATTCAATCCTGCCTGCCTGTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTGAGGTGGGAGCTGGACTGTTGGTTTGATGAGGATTTCAAGATTCATCCTGCCTGCCTGTG 300

Qy    301 AGCTATGCAGTTGCTTTTGTTGCTGGGCCTTGGGCCTTAACCGCCCCAAACCCTATGGCTCTTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCTATGCAGTTGCTTTTGTTGCTGGGCCTTGGGCCTTAACCGCCCCAAACCCTATGGCTCTTC 360

Qy    361 ATCTTCCGCCTCCGACCCTCGGGGATCCAACGGCCACCACCACCTACCACCTGGCATTGTCA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATCTTCCGCCTCCGACCCTCGGGATCCAACGGCCACCACTACCACCTGGCATTGTCA 420

Qy    421 GACACCTTGTATGTCGTCGTCTGCTGCCACCCACCACAACCAC 480
          |||||||||||||||||||||||||||||||||||||||||||
Db    421 GACACCTTGTATGTCGTCGTCTGCTGCCACCCACCACAACCAC 480

Qy    481 TGGGCCCTTTGGCACTGAGATCTGCAAGTCGTCGCCTTTCTTCGTCTTCTTTCTTTCGTGAACCTCTAC 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGGGCCCTTTGGCACTGAGATCTGCAAGTCGTCGCCTTTCTTCGTCTTCTTTCTTTCGTGAACCTCTAC 540

Qy    541 TGCAGTGTCCTTTCCTCACCTGCATCAGCGGTGCACCGCGCTACCTGGGCATCCACCCA 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGCAGTGTCCTTTCCTCACCTGCATCAGCGTGCACCGCGCTACCTGGGCATCTGCACCCA 600

Qy    601 CTTCGGGCACTACGGCTGGGGCGCCCCTCGCCTCAGGCCCTTGGCAGTTTGG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTTCGGGCACTACGGCTGGGGCGCCCCTCGCCTCAGGCCCTTGGCAGTTTGG 660

Qy    661 TTGGTTCGTAGCCGTCGTGCCTCGTGCCAACCTGTTCTTCGTTGTCACCACCAGCAACAAAGGG 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTGGTTCGTAGCCGTCGTGCCTCGTGCCAACCTGTTCTTCGTTGTCACAACCAGCAACAAAGGG 720

Qy    721 ACCACCGTCCTGTTCCCATGACACCACTCGGGCTGAAGAGTTGACACTATGTGCACTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACCACCGTCCTGTTCCCATGACACCACTCGGGCTGAAGAGTTGACACTATGTGCACTTC 780

Qy    781 AGCTCGGCGCGTCATGGGGCTGCTCTTTGGCGTCCTTTGTTTTTGTTATT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGCTCGGCGCGTCATGGGGCTGCTCTTTGGCGTCCTTTGTTTTTGTTTAT 840

Qy    841 GGACTCATGGCTCGTCGCCCTGTATCAGCCCCTTGCCAGGCTCTGCACAGCTGCTCTCGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGACTCATGGCTCGTCGCCCTGTATCAGCCCCTTGCCAGGCTCTGCACAGCTGCTCTCGC 900

Qy    901 CTCCGGCTCCTCTCCGCACCATAGCTGTGGTGGCTGACTGCGTGCTTTCGCCT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CTCCGGCTCCTCTCCGCACCATAGCTGTGGTGGCTGACTGCGTGCTTTCGCCT 960

Qy    961 TTCCACATCACCCGCACCACCATTTACTACCTGGCCAGGCTGTTGGAAGCTGACTGCCGAGTA 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCCACATCACCCGCACCACCATTTACTACCTGGCCAGGCTGTTGGAAGCTGACTGCCGAGTA 1020

Qy   1021 CTGAACATTGTCAACGTGGTCTATAAAGTGACTCGGCCCCTGGCCAGTGCCAACAGCTGC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTGAACATTGTCAACGTGGTCTATAAAGTGACTCGGCCCCTGGCCAGTGCCAACAGCTGC 1080

Qy   1081 CTGGACCCTGTGCTCTACTTCCTCACTGGGAGCGACAAATATCGACGTCAGCTCCGTCAGCTC 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTGGATCCTGTGCTCTACTTCCTCACTGGGAGCGACAAATATCGACGTCAGCTCCGTCAGCTC 1140

Qy   1141 TGTGGTGGTGGCAAGCCCCAGCCCCGCCACGGTCGGTGACTGCTTCTTGCCCCTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGTGGTGGTGGCAAGCCCCAGCCCCGCCACGGTCGGTGACTGCTTCTTGCCCCTG 1200

Qy   1201 CCTGAGGATAGGCAGCTGCAGGTGGGGGGCCCACGGCCCCCCCACCCCCAGGACAGTGAGGG 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCTGAGGATAGGCAGCTGCAGGTGGGGGGCCCACGGCCCCCCCACCCCCAGGACAGTGAGGG 1260

Qy   1261 AGGGCAGATAGATTCTAAACACGGAAGCCCGCAAGTGAGAGAAAAGGGGATGGCAGG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGGGCAGATAGATTGTAAACACGGAAGCCGGCAAGTGAGAGAAAAGGGGATGAGTGCAGG 1320

Qy   1321 GCAGAGGTGAGGGAACCCAAATAGTGAATACCTGGTAAGGTGCTTCTTCCTCTTTTTCCAGGC 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCAGAGGTGAGGGAACCCAAATAGTGAATACCTGGTAAGGTGCTTCTTCCTCTTTTTCCAGGC 1380

Qy   1381 TCTGGAGAGAAGCCCTCACCCTGAGGGTTGCCACGGAGGCAGGGGATATC 1429
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TCTGGAGAGAAGCCCTCACCCTGAGGGTTGCCACGGAGGCAGGGGATATC 1429
```

```
RESULT  3

LOCUS       HSU40223        1651 bp       DNA              PRI       19-JAN-1996
DEFINITION  Human uridine nucleotide receptor (UNR) gene, complete cds.
ACCESSION   U40223
NID         g1117912
VERSION     U40223.1      GI:1117912
KEYWORDS    G protein-coupled receptor; purinoceptor; PCR; intronless; UTP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1651)
AUTHORS     Nguyen,T., Erb,L., Weisman,G.A., Marchese,A., Heng,H.H.,
            Garrad,R.C., George,S.R., Turner,J.T. and O'Dowd,B.F.
TITLE       Cloning, expression, and chromosomal localization of the human
            uridine nucleotide receptor gene
JOURNAL     J. Biol. Chem. 270 (52), 30845-30848 (1995)
MEDLINE     96125054
REFERENCE   2  (bases 1 to 1651)
```

```
AUTHORS   O'Dowd,B.F., Nguyen,T., Marchese,A. and George,S.R.
TITLE     Direct Submission
JOURNAL   Submitted (07-NOV-1995) Brian F. O'Dowd, Department of
          Pharmacology, University of Toronto, 8 Taddle Creek Road, Toronto,
          Ontario M5S 1A8, Canada
FEATURES             Location/Qualifiers
     source          1..1651
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq13"
     gene            391..1488
                     /gene="UNR"
     CDS             391..1488
                     /gene="UNR"
                     /note="introness coding region; Purinoceptor"
                     /codon_start=1
                     /product="uridine nucleotide receptor"
                     /protein_id="AAC50347.1"
                     /db_xref="PID:g1117913"
                     /db_xref="GI:1117913"
                     /translation="MASTESSLLRSLGLSPGPGSSEVELDCWFDEDRKPILLPVSYAV
                     VFYLGLGINAPTLWLFIFRLRPWDPATATYMFHLALSDTIYVSLPTLIYYAAHNHWP
                     FGEICKFVRFLEFWNLYCSVLFITCISVHRYLGICHPLRALRWGRPRLAGLLCLAVW
                     LVVAGCLVPNLFYVTISRKGTYVLCHDTTRPEEFDHYVHFSSAVMGLLFGVPCLVTLV
                     CYGLMARLYQPLPGAAQSSSRLRSLRRTIAVVLIVFAVCFVPFHITRIYILARLLEA
                     DCRVLNIVNVVIKVTRPLASANSCLDPVLYLLTGDKYRRQLRQLCGGKPQPRTAASS
                     LALVSLPEDSSCRWAATPQDSSSCSTPRADRL"
BASE COUNT      290 a    522 c    429 g    410 t
ORIGIN
```

```
Query Match         95.48%;    Score 1363.6;    DB 11;    Length 1651;
Best Local Similarity   98.48%;    Pred. No. 0;    Mismatches 19;    Indels  4;    Gaps  1;
Matches 1389;    Conservative    0;    Mismatches    19;    Indels    4;    Gaps    1;
```

```
Qy      1 AAGGGAGCTTGGGTAGGGGCCAGCCTAGCCTGAGTGCACCCAGATGCGCTTCTGTCAGCT    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 AGGGAGGGCTTGGGTAGGGGCCAGCCTAGCCTGAGTGCACCCAGATGCGCTTCTGTCAGCT   270

Qy     61 CTCCCTAGTGCTTCAACCACTGCTCTCCCTGCTCTACTTTTTTTGCTCCAGCTCAGGGAT   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271 CTCCCTAGTGCTTCAACCACTGCTCTCCCTGCTCTACTTTTTTTTGCTCCAGCTCAGGGAT  330

Qy    121 GGGGGTGGGCGGCAGGGAAAATCCTGCACCCTACTTCTCCCTTCCCATCTCTCCAGGGGGGCC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    331 GGGGGTGGGCGGCAGGGAAAATCCTGCACCCTCACTTCTCCCCTTCCCATCTCCAGGGGGGC  390

Qy    181 ATGGCCAGTACACAGAGTCTCTCCCCTGTTGAGGATCCCTAGGCCTCAGCCCAGGTCCTGGCAGC   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    391 ATGGCCAGTACACAGAGTCTCCCCCGTTGAGATCCCTAGGCCTCAGCCCAGGTCCTGGCAGC  450

Qy    241 AGTGAGGTGGAGCTGGAGCTGGTTGGTTTGATGAGGATTTCAAGTTCATCCTGCTGCCCTGTG   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    451 AGTGAGGTGGAGCTGGAGCTGGTTGGTTTGATGAGGATTTCAAGTTCATCCTGCTGCCCTGTG  510

Qy    301 AGCTATGCAGTTGTGCTTGTTGTTGCTGGCCTTAACGCCCCAAACCCTATGGCTCTTC   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    511 AGCTATGCAGTTGTGCTTGTTGTTGCTGGCCTTAACGCCCCAAACCCTATGGCTCTTC  570

Qy    361 ATCTTCCCGCCCTCCGACCCTGGGATGACAACGGCCACCACCACTACAGTTCCACCTGGTGTCA   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    571 ATCTTCCCGCCCTCCGACCCTGGGATGACAACGGCCACCACCACTACAGTTCCACCTGGTGTCA  630

Qy    421 GACACCTTGATATGTGCTGTCGCGCTGCCCACCCTCAATCACTACTATATGTTCCACCTGGCAGCAC   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    631 GACACCTTGTATGTCGTGTCGCGCTGCCCACCCTCAATCACTACTATATGTTCCACCTGGCAGCAC  690

Qy    481 TGGCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGGCTTTCTTTTCTTTTCTATTGGAACCTCTAC   540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    691 TGGCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGGCTTTCTTTTCTTTTCTATTGGAACCTCTAC  750

Qy    541 TGCAGTGTCCTTTTCCTCACCTGCATCAGCGTGCACCGCTACCTGGGCATCTGGCCACCA   600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    751 TGCAGTGTCCTTTTCCTCACCTGCATCAGCGTGCACCGCTACCTGGGCATCTGGCCACCA  810

Qy    601 CTTCGGGCACTACGGCTGCAGCCACCTAGTGTTCTGCAGGCGCCCAGGCGCCTTCTGCTGCCAGAGTTTGG   660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    811 CTTCGGGCACTACGGCTGCAGCCACCTAGTGTTCTGCAGGCGCCCAGGCGCCTTCTGCTGCCAGAGTTTGG  870

Qy    661 TTGGTGTCTAGCCGCGGTGCCTCGTGCCCAACCTGTTCTTTTGTCACAACACAACAAAAGGG   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    871 TTGGTGTCTAGCCGCGGTGCCTCGTGCCCAACCTGTTCTTTTGTCACAACACAACAAAAGGG  930

Qy    721 ACCACCGGTCCTGTGCCATGACACCAACTCGGCCTGAAGAGTTGACCACTATGTGCACTTC   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    931 ACCACCGGTCCTGTGCCATGACACCAACTCGGCCTGAAGAGTTGACCACTATGTGCACTTC  990

Qy    781 AGCTCGGCGGTCATGGGGGCTGCTCTTTGGGCGTGCGCCTGGTCACTCTTGTTGGCTAT   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    991 AGCTCGGCGGTCATGGGGGCTGCTCTTTGGGCGTGCGCCTGGTCACTCTTGTTTGGCTAT  1050

Qy    841 GGACTCATGGCTCGTGCCTGTATCAGGCCCTTGCCAGGCCTCTGCACAGTGCTTCTCTCGC   900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1051 GGACTCATGGCTCGTGCCTGTATCAGGCCCTTGCCAGGCCTCTGCACAGTGCTTCTCTCGC  1110

Qy    901 CTCCCGCTCCTCCCGACCAACAGCTGTCTGGATCCTGTGCTCTATCTCTTTGCTTCGTTCCT   960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1111 CTCCGATCCTCCCGACCAACAGCTGTCTGGATCCTGTGCTGACTGTCTTTGCTTCGTTCCTG  1170

Qy    961 TTCCACATCACCCGCACCACCATTTACTACCTGGCCCAGGCTGTTGGAGCTGACTGCCGAGTA   1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 TTCCACATCACCCGCACCACCATTTACTACCTGGCCCAGGCTGTTGGAGCTGACTGCCGAGTA  1230

Qy   1021 CTGAACATTGTCAACGTGGTCTATAAAAGTGACTCGGGCCCCTGGGCCTGCCAACAGCTGC   1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1231 CTGAACATTGTCAACGTGGTCTATAAAAGTGACTCGGGCCCCTGGGCCTGCCAACAGCTGC  1290

Qy   1081 CTGGATCCTGTGCTCTACTTGCTCACTGGGGACAAAATATCGACGGACGAGTCAGCTC   1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1291 CTGGATCCTGTGCTCTACTTGCTCACTGGGGACAAAATATCGACGGACGAGTCAGCTC  1350

Qy   1141 TGTGGTGGTGGCAAGCCCCAGCCCGTGCCTGCCTTCTTCCCCTGCGCCACTAGTGTCCCTG   1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1351 TGTGGTGGTGGCAAGCCCCAGCCCGTGCCTGCCTTCTTCCCCTGCGCCACTAGTGTCCCTG  1410

Qy   1201 CCTGAGGATAGCACCTGCAGGTGGGCGGCGGCCACCCCCACCCCCAGGACAGTAGCTGCTCAGTCCT   1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1411 CCTGAGGATAGCACCTGCAGGTGGGCGGCGGCCACCCCCAGGACAGTAGCGTCGCTCTCACTCCT  1470

Qy   1261 AGGGCAGATAGATTCTAAACACGGAAGCCGGCAAGTGAGAGAAAAGGGAATGAGTGCAGG   1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1471 AGGGCAGATAGATTGTAACACGGAAGCCGGCAAGTGAGAGAAAAGGGGAATGAGTGCAGG  1530

Qy   1321 GCAGAGGTGAGGGGAACCCAAATAGTGGTAACCTGGTAGGTGCTTCTTCCCTCTTTTCCC----   1377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1531 GCAGAGGTGAGGGAACCCAAATAGTGGATAACCTGGTAGGTGCTTCTTCCCTCTTTTCCCCAG  1590

Qy   1377 AGGCTCTGGAGAGGAAGCCCTCACCCCTGAGGGT   1377
          |||||||||||||||||||||||||||||| |
Db   1591 GGCTCCTGGAGAGGAAGCCCTCACCCCTGAGGTT   1622
```

```
RESULT 4

HSGPCRE

LOCUS       HSGPCRE     1232 bp    DNA             PRI       11-JUN-1996
DEFINITION  H.sapiens gene encoding G protein coupled receptor.
ACCESSION   X95597
NID         g1296631
VERSION     X95597.1    GI:1296631
KEYWORDS    G-protein coupled receptor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1232)
```

```
AUTHORS     Stam,N.J., Klomp,J., Van de Heuvel,N and Olijve,W.
TITLE       Molecular cloning and characterization of a novel orphan receptor
            (P2P) expressed in human pancreas that shows high structural
            homology to the P2U purinoceptor
JOURNAL     FEBS Lett. 384 (3), 260-264 (1996)
MEDLINE     96197801
REFERENCE   2  (bases 1 to 1232)
AUTHORS     Stam,N.J.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAR-1996) N.J. Stam, NV, Organon, Department of
            Biotechnology and Biochemistry, P.O. Box 20, 5340 BH Oss,
            Netherlands
FEATURES             Location/Qualifiers
     source          1..1232
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="human placenta genomic DNA library"
     CDS             90..1187
                     /codon_start=1
                     /product="G protein coupled receptor"
                     /protein_id="CAA65415.1"
                     /db_xref="PID:e229605"
                     /db_xref="PID:g1296632"
                     /db_xref="GI:1296632"
                     /db_xref="SWISS-PROT:P51582"
                     /translation="MASTESSLLRSGLSLSPGPGSSVEVLDCWDEDEDRFILLPVSYAV
                     VFVLGLGLNAPTLWLFIFRLRPWDATATYMFHLALSDTLYVLSLPLIYYAAHHNWP
                     FGTEICKFVRFLFYWNLYCSVLFLTCISVHRYLGICHPLRALRWGRPRLAGLLCLAVW
                     LVVAGLVPNLFFVVTSNKGTVVTCHDTTREEPFDHYVHFSSAVMGLLFGVPCLVTLV
                     CYGLMARRLYQPLPGSAQSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYILARLLEA
                     DCRVLNIVNVYKVTRPLASANSCLDPVILYLLTGDKYRRQLRQLCGGKPQPRTASS
                     LALVSLPEDSSCRWAATPQDSSCSTPRADRL"
BASE COUNT    204 a    397 c    319 g    312 t
ORIGIN

Query Match       86.0%;  Score 1228.8;  DB 9;  Length 1232;
Best Local Similarity   99.8%;  Pred. No. 2.5e-297;
Matches 1230;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy     92 CTCTACTTTTTTGCTCCAGCTCAGGCGGATGGGGGGTGGGGCAGGAAAATCCTGCCACCCTCA 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCTACTTTTTTGCTCCAGCTCAGGCGGATGGGGGGTGGGGCAGGAAAATCCTGCCACCCTCA 60

Qy    152 CTTCTCCCCTTCCCATCTCCAGGGGGGGGCCATGGCCACTACACAGAGTCCTCCCGTTGAGAT 211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTTCTCCCCTTCCCATCTCCAGGGGGGGGCCATGGCCACTACACAGAGTCCTCCCGTTGAGAT 120

Qy    212 CCCTAGGCCTCAGCCCAGGTCCTGGCAGCAGTGGAGTGGAGCTGGACTGTTGGTTTGATG 271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCTAGGCCTCAGCCCAGGTCCTGGCAGCAGTGGAGTGGAGCTGGACTGTTGGTTGATG 180

Qy    272 AGGATTTCAAGTTCATCCTGCTGCCCTGTGAGCTATGCAGTTCAGTGTTGTCTTTGTTG 331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGGATTTCAAGTTCATCCTGCTGCCCTGTGAGCTATGCAGTTCAGTGTTGTCTTTGTTG 240

Qy    332 GCCTTAAACGCCCCAACCTARGGCTCTTCACTTCATCTTCCGCCTCGGGATTGCAACGG 391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCCTTAAACGCCCCAACCTATGGCTCTTCATCTTCCGCCTCCGGGATGCAACGG 300

Qy    392 CCACCTACAGTTCCACCCTGGCATTGTGCAGACACCTTGTATGTGCTTCGGCTGCCCACCC 451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCACCTACAGTTCCACCCTGGCATTGTGCAGACACCTTGTATGTGCTTCGGCTGCCCACCC 360

Qy    452 TCATCTACTATTATGTGCAGCCCACAACCACTGGCCCCTTTGGCACTGAGATCTGCAAGTTCG 511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCATCTACTATTATGTGCAGCCCACAACCACTGGCCCCTTTGGCACTGAGATCTGCAAGTTCG 420

Qy    512 TCCGGCTTTCTTTTCTATTGGAACCTCTACTGCAGTGTCCTTTTCCTCACCTGCATCAGCG 571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCCGGCTTTCTTTTCTATTGGAACCTCTACTGCAGTGTCCTTTTCCTCACCTGCATCAGCG 480

Qy    572 TGCACCGCTACCTGGGCATCTGCCACCACTTCGGGCACTACGCTGGGGGCGCCCCTCGCC 631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCACCGCTACCTGGGCATCTGCCACCACTTCGGGCACTACGCTGGGGGCGCCCCTCGCC 540

Qy    632 TCGCAGGCCCTTCTCTGCCCTGGCCAGTTGGTTGGGGTCGGGGTAGGGGTGGGGTGCCCAACC 691
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCGCAGGCCCTTCTCTGCCCTGGCCAGTTGGTTGGGGTCGGGGTAGGGGTGGGGTGCCCAACC 600

Qy    692 TGTTCTTTGTTCACAACCAGCAAACAAAAAGGGAGAAGGACCACGTGTCTGTTGCCATGACACC 751
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGTTCTTTGTTCACAACCAGCAAACAAAAAGGGAGAAGGACCACGTGTCTGTTGCCATGACACC 660

Qy    752 CTGAAGAGTTTGACCACTTCAGCTCTCGGCACTGCCGGTCATGGGGGCTGCCTCTTTGGCG 811
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CTGAAGAGTTTGACCACTTCAGCTCTCGGCACTGCCGGTCATGGGGGCTGCCTCTTTGGCG 720

Qy    812 TGCCCTGCCTGGTCACTCTTGTTTGGCTATGGACTCATGGGGCGGCCCTGTATCAGCGCCCT 871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TGCCCTGCCTGGTCACTCTTGTTTGGCTATGGACTCATGGGGCGGCCCTGTATCAGCGCCCT 780

Qy    872 TGCCAGGGCTCTGCACACAGTGCGTCTTCTCGCCTCTCCCGCCACCATAGCTGTGGTGC 931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGCCAGGGCTCTGCACAGTGCGTCTTCTCGCCTCTCCCGCCACCATAGCTGTGGTGC 840

Qy    932 TGACTGTCGTTTGCTTTGGTGTCTGCTTTGGTGCCTTTCACATAACCCGGACCATTACCTGG 991
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TGACTGTCGTTTGCTTTGGTGTCTGCTTTGGTGCCTTTCACATAACCCGGACCATTACCTGG 900

Qy    992 CCAGGCTGTTGGAAGCTGCGAGTACTGAACATTGTCAACGTGTGGTTCTATAAAGTGA 1051
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCAGGCTGTTGGAAGCTGCGAGTACTGAACATTGTCAACGTGTGGTTCTATAAAGTGA 960

Qy   1052 CTCGGGCCCCTGGCCAGTGCCAACAGCTGCCCTGGATCCCTGTGCTCTACTTGCTCACTGG 1111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CTCGGGCCCCTGGCCAGTGCCAACAGCTGCCCTGGATCCCTGTGCTCTACTTGCTCACTGG 1020

Qy   1112 ACAAATATCGACGTCAGTCCGTCCAGCTCTGTGGTGGGTGGCAAGCCCCAGCCCCGCACGG 1171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACAAATATCGACGTCAGTCCGTCCAGCTCTGTGGTGGGTGGCAAGCCCCAGCCCCGCACGG 1080

Qy   1172 CTGCCTCTTCCCTGGCACTAGTGTCCCTGCCTGAGGATAGCAGCTGCAGTGGGCGGGGCA 1231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTGCCTCTTCCCTGGCACTAGTGTCCCTGCCTGAGGATAGCAGCTGCAGTGGGCGGGGCA 1140

Qy   1232 CCCCCCAGGACAGTAGCTGCTCTACTCCTAGGGCGGCAGATAGATTCTAACACGGGAGCCGG 1291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CCCCCCAGGACAGTAGCTGCTCTACTCCTAGGGCAGATAGATTCTAACACGGGAGCCGG 1200

Qy   1292 CAAGTGAGAGAAAAGGGGATGAGTGCAGGGCA 1323
          ||||||||||||||||||||||||||||||||
Db   1201 GAAGTGAGAGAGAAAAGGGGATGAGTGCAGGGCA 1232

RESULT 5
LOCUS       RNP2Y4RG      2684 bp    DNA             ROD       03-SEP-1997
DEFINITION  Rattus norvegicus P2y4 receptor gene.
ACCESSION   Y14705
NID         g2370438
VERSION     Y14705.1  GI:2370438
KEYWORDS    G-protein coupled receptor; P2Y4 gene.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 2684)
AUTHORS     Bogdanov,Y.D., Wildman,S., King,B.F. and Burnock,G.
TITLE       Molecular cloning and genomic structure of rat P2y4 receptor
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 2684)
AUTHORS     Bogdanov,Y.D.
TITLE       Direct Submission
JOURNAL     Submitted (27-AUG-1997) Y.D. Bogdanov, University College London,
            Dept. of Anatomy and Dev. Biology, Gower Street, London WC1E 6BT,
```

FEATURES   Location/Qualifiers
    source   1..2684
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /dev_stage="adult"
                     /clone="4-1"
                     /tissue_type="liver"
    gene   1358..2443
                     /gene="P2Y4"
    CDS   1358..2443
                     /gene="P2Y4"
                     /codon_start=1
                     /product="G-protein coupled receptor"
                     /protein_id="CAA75007.1"
                     /db_xref="PID:e339607"
                     /db_xref="PID:g2370439"
                     /db_xref="GI:2370439"
                     /db_xref="SPTREMBL:O35811"
                     /translation="MISAESLLFTSLGPSPSSGGDGCRFNEEFKFILLPMSYAVVFVL
                     GLALNAPTLWLFLFRLRPWDARPATYMFHLALSDTLYVLSLPTLVYYAARNHWPFGTG
                     LCKFVRRLFYWNLYCSVLFLTCISVHRYLGICHPLRAIRWGRPRFASLLCLGVWLVVA
                     GCLVPNLFFVTNANGTTILCHDTTLPEEFDHYVYFSSAVMVLLFGLPFLIILVCYGL
                     MARRLRPLPGAGQSSSRLRLRTIAVVLTVFAVCFVPFHITRTIYYQARLLQADCHV
                     LNIVNVYKVRPLASANSCLDPVLIYLFTGDKYRNQLQOLCRGSKPKPRTAASSLALV
                     TLHEESISRWADTHQDSTFSAYEGDRL"

BASE COUNT   604 a   662 c   669 g   749 t
ORIGIN

Query Match   56.1%;   Score 802.2;   DB 12;   Length 2684;
Best Local Similarity   81.3%;
Matches   963;   Conservative   0;   Mismatches 203;   Indels   19;   Gaps   2;

Qy   174 GGGGGCCATGGCCAGTACAGAGAGTCCTCCCGTTGAGAGATCCCTAGGGCCTTCAGCCCAGGGTCC 233
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1351 GGGGGCCATGACCAGTCCAGAAATCCTTGGCTATTCACATCACTACTGGAGTTCCCAGAGTTC 1410

Qy   234 TGGCACGAGTGAGGTGGAGCTGGACTGTTGGGTTTGATGAGGGGAATTCAAGTTCATTCCTGCT 293
          ||||| |||||||||||||||||||||||||||||||||||||
Db   1411 TGGAGATGGGTGAT-----------TGTAGGTTTAATCAGAGAGTTCAAGTTCATCCTGTT 1458

Qy   294 GCCTGTGAGCTATATGCACGTTGTCTTTGTGGCTCGGGGCCTTAACGCCCCAACCCTATG 353
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1459 GCCTATGAGCGTATGCAGTTGGTGTTTGTGCCTGGGCCCCTCAATCGTCCAACCCTCTG 1518

Qy   354 GCTCTTCATCTTCCGCCTCCGACCCTGGGATGCAACGCCCACCTACACACTGGCC 413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1519 GCTCGTTCCTCTTCCGCCTTCGACCCTGGGATCCAACAGCCACCTACACTGGGC 1578

Qy   414 ATTGTCAGACACCTTGTATGTGTCGCGTCGCGGGCCCACCCTCATCTCACTACTATTATGCAGCCCA 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1579 ATTGTCAGACACCTTGTATGTGTCGTCACTGCCCACCCTCGTTCTACTACTACTGCCAG 1638

Qy   474 CAACCACTGGGCCCCTTTGGCATCGCAAGTCGTCCGCGCTTCTCTTTTCTATTGGAA 533
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1639 AAACCACTGGGCCCTTTGGCACTGGCCTTTGCAAGTTTGTCCGCGCTTTCTCTTTCTATTGGAA 1698

Qy   534 CCTCTACTGCAGTGTCCTTTTTCCTCACCTGCAACAGGGTGCACGGCTACCTGGGCACTG 593
          |||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1699 CCTCTACTGTAGTGTCCTTTTCCTCACCTGCAATCAGTGTGCACCGATACCTGGGTATCTG 1758

Qy   594 CCACCCACTTCGGGCACTACCCTGGGGGCGCCCCTCCCGCCCCGCCAGGCCTTCTGCCTGGC 653
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1759 CCACCCACTGCGGGCAATCCGCTGGGGCGCCCCTCGATTTGCAAGCTTCTCGCTGGGG 1818

Qy   654 AGTTGGTTGGTTCGTAGCCGGCGCGGCTCCGTGCCCACCACCGTTCTTGTTCTTGTCACAAACCAGCAA 713
          ||||||| ||||||||||||||||||||||||||||||||||||||
Db   1819 TGTTGGTTGGTTAGTAGCTGGCTGGCTGCCCCTGTGCCCGCTGCTCATTATCCTACTATGT 1878

Qy   714 CAAAGGGACCACCGTCCTGTGCCATGACACACCACTCGGCCTGAGAAGAGTTTGACCACTATGT 773
          || ||| ||||||||||||||||||||||||||||||||||||||||||
Db   1819 TGTTGGTTGGTTAGTAGCTGGCTGGCTGCCCCTGTGCCCGCTGCTCATTATCCTACTATGT 1878

LOCUS   RNP2Y4REC   1177 bp   mRNA   ROD   17-SEP-1998
DEFINITION R.norvegicus mRNA for G-protein linked receptor P2Y4.
ACCESSION   Y11433
NID   g3618228
VERSION   Y11433.1 GI:3618228
KEYWORDS   G protein-linked receptor; P2Y4 gene; P2Y4 receptor.
SOURCE   Norway rat.
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 1177)
  AUTHORS   Webb,T.E., Henderson,D., Roberts,J.A. and Barnard,E.A.
  TITLE   Molecular cloning and characterisation of the rat P2Y4 receptor
  JOURNAL   J. Neurochem. 71, 1424-1434 (1998)
REFERENCE   2 (bases 1 to 1177)
  AUTHORS   Webb,T.E.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-FEB-1997) T.E. Webb, Royal Free Hospital School of
            Medicine, Molecular Neurobiology Unit, Rowland Hill Street,
            Hampstead, London NW3 2PF, UK
FEATURES   Location/Qualifiers
     source   1..1177
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="brain"
                     /clone="rP2Y4.pBK/CMV"
     gene   55..1140

RESULT   6
RNP2Y4REC

Qy   774 GCACTTCAGCTCGGCGGTCATGGGGCTGCTCGTTCTTTGGCGTGCCCTCTGCCAGAGGAGAGTTTGACCACTTTGT 833
          ||||||||| ||||||||||||||||| |||||| ||||||||| || || |||||||
Db   1879 CTACTTCAGTTCGGCAGTCATGGTGCTGCTCTTCGGTCTTCCCTTCCTTGTTATCACCCTGGT 1938

Qy   834 TTGCTATGGACTCATGGCTCGTGGCCCTGTATCAGCCCCCTTGCCAGGCGTCGTC 893
          |||||||||||||||| || ||| |||| ||||||||||||||||||||
Db   1939 CTACGGACTCATGGCCCGGCGACTGTATCCGACCTTTGCCAGGAGTGTGGACAGTCATC 1998

Qy   894 TTCTCGCCCTCCGCTCTCCGCACCATGAGCTGTGGTGGTGGTGGTGGTTTGCTTTGCTCTT 953
          ||||| ||||| || || ||||||| |||||||||||||||||||||||||||
Db   1999 CTGCTCTATGGACTCATGGCCCGGCGACTGTATCCGACCTTTGCCAGGAGTGTCATC 2058

Qy   954 CGTGCCCTTTCCACATCACCCGCACCAATTTACTACCTGCCAGGCGTGTTGGAAGCTGACTG 1013
          ||||| || ||||| ||||||||||||||||||||||||||||||||||||
Db   2059 TTCTCGGGCTCCGTTCCCTCCGTCACCATAGCCATTGCTGTGGTTCTGACTGTGTCTGCTT 2118

Qy   1014 CCAGTAGTCTGAACATTGTCAAGCGTGGTCTCTATAAAGTGACTCGGCGCGCCCCTGGCCAGTGCCAA 1073
          || ||||||||||||||||||||| |||||||||||||||||||||||||
Db   2119 CGTGGCCTTTCCACATCACCCGCACAATTTATTACCAGGCAAGACTGTTGCAAGCTGACTG 2178

Qy   1074 CAGCTGCCTGGGATCCTGTGCCTCTACTTGCCTCACTGGGGACAAAATATCGACGTCAGCCTCG 1133
          || ||||||| |||| ||||| || || ||||| ||| ||||||| ||||||
Db   2179 TCAAGTGCTGAACATTGTCAAGTGGTTTACAGGTGACACCTCACTGGGTATCTG 2238

Qy   1134 TCAGCTCTGTGGTTGGTGGTGGCAAGCCCCACCCCCACCCCCCACGGGCGTGCCTCTTCCCTGCCGACTAGT 1193
          |||||||||||| ||||||||| ||||||||||||||||||||||||||||||||||
Db   2239 TAGCTGTCTTGATCCAGTCGTCCTCATCTCTTCACGGGACAAGTATCCGAAACCAGCTCCA 2298

Qy   1194 GTCCCTGCCTGAGGATAGCAGCTGCAGGTGGTGGCCACCCCCAGGACAGCAGTAGCTGCTC 1253
          ||||| |||||||||||| ||| |||| ||||||||| ||||||||||||||||
Db   2299 GCAGCTATGCCGAGGCCAGCGACAAAACCCAAGCCCGTGCTCCTCCCTCCTTCTCTGCCTTGGT 2358

Qy   1254 TACTCCCTAGGGCAGATAGAATTCTAAACACGGAGGAAGCCGGCAAGTGAGAGAAAAGGGGATGA 1313
          | || || |||||| | ||||||||| ||| || ||| ||||||||||||||||||||
Db   2359 GACTTCACATGAGGAAAGCCATCAGAGACAGGTGACGGCGGGGACATTCACGCCGACTGGGGGAAGCTGA 2418

Qy   1314 GTGCAGGGGCCAGAGGGTGAGGGAACCCAATAAGTGATAATACCTGGGTAAGG 1358
          ||||| | ||||| ||||| ||||||| |||
Db   2419 TGCTTATGAGGGAGGGATCGGATTTGTAAC------CCTGCTCGGGATATGAGAGAGGAGGAACTAA 2471

Qy   1358 CTGCAGGGGCCAGAGGGGGAGGGAGGGAGCTTGTTGTTAGTATCACTTGACACGGG 2516
          |||||||||||| ||||||||||||||||||||||
Db   2472 CTGCAGGGGCCAGAGGGGGAGGGAGGGAGCTTGTTGTTAGTATCACTTGACACGGG 2516

gene

```
CDS             /gene="P2Y4"
                55..1140
                /gene="P2Y4"
                /codon_start=1
                /product="G protein-linked P2Y4 receptor"
                /protein_id="CAA72241.1"
                /db_xref="PID:e305322"
                /db_xref="PID:9361829"
                /db_xref="GI:3618229"
                /db_xref="SPTREMBL:O35811"
                /translation="MTSAESLLFTSLGPSPSSGGDGCRNEEFKFILLPMSYAVVFVL
                GLAINAPTLWLFLFRLRPWDATATYMFHLALSDTIVLSLPLTVYYYAARNHWPFGTG
                LCKFVRFLFYNNLYCSVLFLTCISVHRYIGICHPLRAIRWGRPRPASLLCLGVWLVVA
                GCLVPNLFFVTTNANGTTLICHDTTLPEEFDHYVFSSAVMVLLFGLPFLITLVCYGL
                MARLYRPLPGAGQSSRLRSLRTIAVVLTVFAVCFVPFHITRTIYYQARLLQADCHV
                LNIVNVVKVTRPLASANSCLDPVLYLFTGDKYRNQLOQLCRGSKPKPTAASSLALV
                TLHEESISRWADTHQDSTFSAYEGDRL"
```

```
BASE COUNT      224 a      348 c      283 g      322 t
ORIGIN
```

```
Qy     174 GGGGGCCATGGCCAGTACAGAGTCGCTCTCCCGTTGAGAATCCCTAGGCCTCAGCCCAGGTCC 233
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48 GGGGGCCATGGCCAGTGCAGAAATCCTTGCTARTTCACATCACTAGGTTCCAGCCCAAGTTC 107

Qy     234 TGGGCAGCAGTCAGGTGGAGCTGGGACTGTTGGTTGTTGAATGAGGATTCAAGTTCATCCTGCT 293
           ||||||||||||||| |||                        ||||||||||||||||| |||
Db     108 TGGAGAATGGTGAT--------------------TGTAGGTTTAATGAGAGTGCAAGTTCATCCTGTT 155

Qy     294 GCCGTGTGAGCTATGCAGTTGCTCTTTGTGTCTGGGCCTTAACGCCCCAACCCTATG 353
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     156 GCCTATGAGCTATGCAGTTGTGTTTTGCTGGGCCTTCAATGCTCAAACCCTCTG 215

Qy     354 GCTCTTCATCTTTCCGGCCTCCGACCCTGGGATGAGCAAGGGCCACCTACATGTTCCACCTGGC 413
           ||| ||| ||||  ||||||||||||||||||||||||||||||||||||||||||||
Db     216 GCTCGTTCCCTTCCGCCCTTCCGGATGCAAGACGCCACCTACAGTGTTCCACCTGGC 275

Qy     414 ATTGTCAGACACCTTGTATGTGCTGCTGCCCTGCCCACCCTCCCCACCCTCATTCTATTATGCACGCCCA 473
           ||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db     276 ATTGTCAGACACCTTGTATGTGCTGTCACTCCCACCCCTCGTCTACTACTAGCTGCCAG 335

Qy     474 CAACCACTGGGCCCCTTGGCCACTGAGAATCTGCAAGTGTCCCGCCTTTCTTTTTCTTATTGGAA 533
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     336 AAACCACTGGGCCCTTTGGCCACTGAGTTGTTGTCCGGCTTTCTTCTTCTTATTGGAA 395

Qy     534 CCTCTACTGCAGTGTGCCTTTTCCTCACCTGCATCAGCCGCTGCACCGCTGTTCTTTGGGGCACTCTG 593
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     396 CCTCTACTGTGTAGTGTCCTTTTCCTCACCTGCATCAGTGTGCACCGATACCTGGTATCTG 455

Qy     594 CCACCCACTTCGGGGCACTACCTGCGGGGCGCCCCGCCCGATTGCAAGGCCTTCTTCTTTGCCTGGC 653
           ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     456 CCACCCACTCGGGCAAATCCGGCTGCTGGGGCGCGCGATTGCAAGGCCTTCTTCGTGCCTGGG 515

Qy     654 AGTTTGGTTGGTTCGTAGCCGGCGCGCTGCCCGTGTCGTGCCGCCAACCAGCAA 713
           |||||||| ||| |||||| |||||||||||| |||||||||||||||||||||||
Db     516 TGTTTGGTTGGTAGTAGCTGGCGCGCTCGCCGTCGCCAATCTTTCTTTGTGACAAGCAATGC 575

Qy     714 CAAAGGGACGACACCGTCGTGTCCATGAGCAACACCACTCGGCCTCGAAGAGTTGACCTATATGT 773
           ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     576 CAATGGAAGCTACCATCCGTGCCATGAGCACTCTCCAGGAGAGTTTGACCACTACGT 635

Qy     774 GCACTTCAGCTCGGGCGTCATGGGGGCTGGCCGTGCCCTGCCGTGCACTCTTGT 833
           ||| ||||||||||||| |||||| |||||||||||| |||||||||||||||
Db     636 CTACTTCAGTTCGGCAGTCATGGTGCTGCTCTTTGGTTTGCCCTTCTTGATCACCCTTGT 695

Qy     834 TTGCTATGGACTCATGGCCTCGTCGCCTGTATCAGCCCTGTGCTC 893
           ||| |||||||||||||||||||||| || ||||||||||||||
Db     696 CTGCTATGGACTCATGGCCCGGCGACTGTATCGACACCTTGCCAGGAGCTGACAGTGCATC 755

Qy    894 TTCTCGCCTCCGCTCTCTCCGCCATAGCTGTTGGTGGTTGCTGAGTGTCTGCTTCTTTTTGCTGTCTTGCTT 953
          |||||||                  ||||||||||||||||| |||||||| ||| |||||
Db    756 TTCTCGGCTCCGCTCCCTCCGTTACCAATGCTGGTGGTTCTTGAGTGTTCTTCTTTTGGTTTT 815

Qy    954 CGTGCCTTTCCACATACACCCGCGACCAATTTACTTACTGGTACGGCCAGGCTGTTGGAAGCTGACTG 1013
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    816 CGTGCCCTTCCACATCACCCGCACAATTTATTACCAAGCAAGACTGTTGCAAGCTGACTG 875

Qy   1014 CCGAGTACTGAACATTGTCAACGTGGTCTATAAAGTGACTCGGCCCCTGGCCAGTGCAA 1073
          |||||||||||||||||| ||||| ||| |||||||||||||||||||||||||||||||
Db    876 TCATGTGCTGAACATTGTCAARGTGGTTTTACAAGGTGACTCGACCCCTGGCCAGTGCTAA 935

Qy   1074 CAGCTGCCCTGGATCCGTGCTCTACTTGCTCACTGGCGACAAATATCGACGGTCAGCTCCG 1133
          |||||| |||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    936 TAGCTGTCTTGATCCAGTCGTTCTTCCTGCTCACGGGGACAAGTATGAAAACCAGCTCCA 995

Qy   1134 TCAGCTCTGTCGTGGTGGGTGGGCAAGCCCCCAGCCCCAGCCCAACGCCAACCAGCTGCC 1193
          ||| | ||||||||||||||||||||||||||||||||||||||||||| ||||||||| ||
Db    996 GCAGCTARTGCCGGAGGCAGGCAAAACCAAGCCCAAGCCGACGGCTGCCTCTTCCCTGGT 1055

Qy   1194 GTCCCTGCCTGAGGATAGCAGCTGCAGGTGGGGCGCCACCCCCCAGGACAGTAGCTGCTC 1253
          |||||||||||||||||||||| ||||||||||||| |||||||||||||||||||||||
Db   1056 GACTCTACATGAGGAGAAGCATCAGCAGCAGGTGGGGCGGGACAACCACCAAGAACAGCACTTCTC 1115

Qy   1254 TACTCCCTAGGGCCAGATAGAATCTAACACGG 1283
          |||||||| |||||||||||||||||||||||
Db   1116 TGCTTATGAGGGAGAGATCGAATGTNAACCCTG 1145
```

```
RESULT 7
AF031897
LOCUS      AF031897              1977 bp    mRNA            VRT       06-JAN-1998
DEFINITION Meleagris gallopavo G protein coupled P2Y nucleotide receptor
           (tp2y) mRNA, complete cds.
ACCESSION  AF031897
NID        g2707255
VERSION    AF031897.1  GI:2707255
KEYWORDS   .
SOURCE     common turkey.
  ORGANISM Meleagris gallopavo
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
           Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.
REFERENCE  1  (bases 1 to 1977)
  AUTHORS  Boyer,J.L., Waldo,G.L. and Harden,T.K.
  TITLE    Molecular cloning and expression of an avian G protein-coupled P2Y
           receptor
  JOURNAL  Mol. Pharmacol. 52 (6), 928-934 (1997)
  MEDLINE  98086419
REFERENCE  2  (bases 1 to 1977)
  AUTHORS  Boyer,J.L., Waldo,G.L. and Harden,T.K.
  TITLE    Direct Submission
  JOURNAL  Submitted (29-OCT-1997) Dept. of Pharmacology, University of North
           Carolina School of Medicine, CB# 7365 Mary Ellen Jones Building,
           Chapel Hill, NC 27599-7365, USA
FEATURES             Location/Qualifiers
     source          1..1977
                     /organism="Meleagris gallopavo"
                     /db_xref="taxon:9103"
                     /tissue_type="blood"
     gene            1..1977
                     /gene="tp2y"
     CDS             374..1498
                     /gene="tp2y"
                     /codon_start=1
                     /product="G protein coupled P2Y nucleotide receptor"
                     /protein_id="AAC60339.1"
                     /db_xref="PID:g2707256"
                     /db_xref="GI:2707256"
                     /translation="MDAPVRMFSLAPWTPTPFWLGGNTTAAAAEACKVFNEEFKFILL
                     PISYGIVFVVGLPLNSWAMWIFVSRMRPWNARTTYMFNLAISDTLYFSLLPTLVYYA
                     DRNNWPFGKVFCKIVREFLFYANLYSSILFLTCISVHRYMGICHPIRSLKWVTKHARL
                     ICVGVWLVVTLCLIPNLIFVTTSSKDNSTLCHDTTKPEEFDHYVHYSSSSIMALLFGIP
                     FLVIVVCYCLMARLCKRSFPSPSRVPSYKKRSIRMIIIVLTVFAICFVPFHITRTL"
```

```
Qy   1179 TTCCCTGGCACTAGTGTCCCTGCCTGAGGATAGCAG 1214
          |||||  ||||||| |||||||| |  /  ||||||||
Db   1393 CCTGCTGGCCTCGTGGTCCCCCTCCGTGGACAGCAG 1428
```

YYTSRYFQADCQTLNINIINFTYKIIRPLASINSCLDPILLFMAGDKYRGRLIRRGAAQRP
RPVPTSLLALVSPSVDSSVVGSCCNSESRGMGTVWSRGGQ"

```
BASE COUNT    416 a    607 c    531 g    423 t
ORIGIN
```

Query Match   Best Local Similarity 29.1%; Conservative 65.3%; Pred. No. 3.8e-94; DB 4; Length 1977;
Matches 650; Conservative 0: Mismatches 331; Indels 15; Gaps 2;   Score 416.4;

```
Qy    228 AGGTCCTGGCAGCACAGTGAGGTGGAGCTGGAGTGGATTTCAAGTTCAT 287
          |||                                         ||||||
Db    439 AGGGAATACAAACCCCTGCTGCAGAGGCAAAAGTGGGTCTTCAACGAGGAGTTCAAGTTCAT 498

Qy    288 CCTGCTGCCTGTGAGCTATGCCAGTTGTCTTTTGTCTGGGCTTGGGCCTTAACGCCCCAAC 347
          |||                  ||||  |  | |   | ||||  | ||||||  |||||| ||
Db    499 CCTGCTGCCATCCTCTACGGCATCGTGTCTTCGTTGGTCTTCGTCGGGTTGCCCCTCAACTCCTGGGC 558

Qy    348 CCTATGGCCTCTTCACTCTTCCGCCCTCAGAGGTGGATGGAAACGGCCACCTACACGTTCCA 407
          | |   ||||| |   |  |||| |||||||  || | |  |  | ||| |||| ||| |||
Db    559 CATGTGGATTTTGTCTCCAGGAATGAGCCCTTGGAATGCCACCACTACTTACACGTTCAA 618

Qy    408 CCTGGCATGTGCAGACACCTTGTATGTGCTGGCTGTCCGGCGTCCACCCTCATCTACTATTATGC 467
          || ||   |   | |    ||| |  | ||||   |  ||  |  |||  | |||||| ||||| |
Db    619 CCTGGCTACTCTGACACTCTATGTCTTCTCCCCCAACCCTGGTCTAACCCTGGTCTTCA 678

Qy    468 AGCCCACAACCACTGGCCCTTTGGCACTGAGATCGTTCGTCCGGCTTTCTTTCTA 527
          |  | | | ||| ||| || | ||||  | | || ||  ||  ||| |||||  |
Db    679 TGACCGCAACAACTGGCCCTTCGGGGAAAGTGTTCTGCAAGATCGTGCGCTTCCCTCTTCTA 738

Qy    528 TTGGAACCTCTACTGCCAGTGTCCAGTGGTCGTAGCGCTGGG 587
          || |   |||| |||   |||||||  | || |||||   |||
Db    739 CGCCAACCTCTACAGCAGCGATCCTTCCTGACGTGCATCAGCGTGCCCACCGCTACGTGGG 798

Qy    588 CATCTGCCACCCACTTCGGGGCACTACGCTGGGGCCCCCTTCTCTG 647
          |||||||||||| |   |||    ||  |  ||||| |||||| |||
Db    799 CAATCTGCCACCCATCCCTCAAGTGGGTGAAGACCAAAACATGCACGACTCATCTG 858

Qy    648 CCTGGCAGTTTGGTTGGTCGTAGCGGCTCCCTCGTGCCCAACCGTTCTTTGTCACAAC 707
          || ||||  | | || ||  | || |||||| |   ||| ||||| || | |||||||
Db    859 CGTGGGGCGTCTGGCCTGGTTGTCACCATCTGCCTCATCAGCGCCTCAGCCTACATGGG 918

Qy    708 CACCACAAAAGGAGACCACACGGTCGTGCATGACACCACTTACCCCGGCCTGAAGAGTTTGACCA 767
          |||                  |   |||||  |||  | || |||| || ||||||||| |||||
Db    919 CAGCTCCAAGGACAACAGACACCCTTTGCCATGACACCAAAACCTGAGGAGTTCGACCA 978

Qy    768 CTATGTGCACTTCAGCTCGGCGGGTCATGGGGCGTCTCTTTGGCGTGCCCTGGGTCAC 827
          |   | | |||||| |  ||||||||| || | |||  ||  || ||| ||| |  |
Db    979 CTACGTGCGACTACAGCTCCTTCAGCAGTGGCGCTCCTCTTCGGTATTCCCTTCCTCGGTGAT 1038

Qy    828 TCTTGTTTCCTATGGACTCATGGCTCGTGCCCTGTATCAG---------- 878
          ||| ||  | ||  |    | ||| | | |||| | | ||  |||| || ||
Db   1039 TGTTGTGTGTTCTACTGCCTGATGGGCCAAGAGGCTCAGGAGCTTCCCAGCCAG 1098

Qy    879 CTCTGCACAGTGTCGTCTTCTCTGCGGCTCCGCGCTCCGCGCACCAGTGTGT 938
          || |  | |||||| |  | |||   | || |   |||| | |||| |||||
Db   1099 CCCCGCGTGCCTCCCTACAAGAAGCGCTCCATCAAGATGATCATCGTGCTCACCGT 1158

Qy    939 CTTTGCTGCTGCTTTCGGTGCTGCCTTTCCACATCACCCGCACCCATTTACTTACCTGCCAGCT 998
          | |||| |  |      | |||| ||  |||  ||| |   ||| |   || | |  | |||
Db   1159 CTTCGCCATTCGCTTTGTGGTGCCTTCCACACCCTACTCCTTCCTGGGCGAT 1218

Qy    999 GTTGGAAGCTGACTGCGCCAGTACTGACGAACATTGTCAACGTGGTCTTATAAAGTGACTCGGCC 1058
          | |||| | || | |||| |||| ||| ||  | || | || |    ||  ||| | | |  |||
Db   1219 CTTCCAAGCTGATTGTCAGACCCTGACAACACTTCACTGACAGATCACGCGACC 1278

Qy   1059 CCTGGGCCAGTGCCGTCAAACACGTGCTGCTGGATCCGTGTCTCTTACTGCTTCAGGGGACAAATA 1118
          || |||||| || || | || | || || |||| |||||| | |   |||||| |  ||| || |
Db   1279 CCTGGGCCAGCATCAAACAGCTGCCCTGGACCCCATCCTGTACTTCATGGCTGGGGACAAGTA 1338

Qy   1119 TCGAACGTCAGCTCCGTCAGCTCTTGTGGTGGTGGCAAAGCCCCAGCCCGCACGGGCTGCCTC 1178
          | ||||||||||||||  | ||||||   | || |    | |||  ||| | |   | || |
Db   1339 CCGGGGACGGGCT-----GCGCGGGGGCAGCCCAGCGGGGCCCGTGCCCAACCCTATGGCTGTCTTCAT 1392

Qy   1179 TGAAGGTTAGCCATTTGCTTTGTTCGTCTGCCAG 1392
          |  |||| |||||  | |||||  |  ||  |  |
Db   1393
```

```
RESULT    8
LOCUS       HSU07225     2025 bp    mRNA                    PRI      05-MAR-1998
DEFINITION  Human P2U nucleotide receptor mRNA, complete cds.
ACCESSION   U07225 S74902
NID         g984506
VERSION     U07225.1  GI:984506
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2025)
  AUTHORS   Parr,C.E., Sullivan,D.M., Paradiso,A.M., Lazarowski,E.R.,
            Burch,L.H., Olsen,J.C., Erb,L., Weisman,G.A., Boucher,R.C. and
            Turner,J.T.
  TITLE     Cloning and expression of a human P2U nucleotide receptor, a target
            for cystic fibrosis pharmacotherapy [published erratum appears in
            Proc Natl Acad Sci U S A 1994 Dec 20;91(26):13067]
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (8), 3275-3279 (1994)
  MEDLINE   94211846
REFERENCE   2  (bases 1 to 2025)
  AUTHORS   Parr,C.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-1994) Claude E. Parr, University of North
            Carolina, Dept. of Medicine, Pulmonary Diseases, 724 Burnett-Womack
            Bldg., Chapel Hill, NC 27599, USA
COMMENT     On Sep 13, 1995 this sequence version replaced gi:487152.
FEATURES             Location/Qualifiers
     source          1..2025
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="5A2116"
                     /clone_lib="CF/T43"
                     /cell_line="CF/T43"
                     /cell_type="epithelial cell"
                     /tissue_type="airway epithelium"
     5'UTR           1..245
     CDS             246..1379
                     /note="P2U receptor; similar to mouse ATP receptor,
                     GenBank Accession Number L14751"
                     /codon_start=1
                     /function="G protein-coupled surface membrane receptor"
                     /product="P2U nucleotide receptor"
                     /protein_id="AAC04923.1"
                     /db_xref="PID:g984507"
                     /db_xref="GI:984507"
                     /translation="MAADLGPWNDTINGTWDGDELGYRCRFNEDFKYVLLPVSYGVVC
                     VLGLCLNAVALYIFLCRLKTWNASTTYMFHLAVSDALYAASLPLLVIYARGDHWPFS
                     TVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPLRSLRWGRARYARRVAGAVVWL
                     VLACOAPVLYFVTTSARGVRTCHDTSAPELFSRFVAISSVMLGLLFAVPFAVILVCY
                     VLMARRLLKPAYGTSGGLPRAKRKSVRTIAVVLAVFALCFLPFHVTRTLYYSRSLDL
                     SCHTLNAINMAYKVTRPLASANSCLDPVLYFLAGQRLVREFARDAKPPTGPSPAFPARR
                     RLGLRRSDRTDMQRIGDVLGSSEDFRRIESTPAGSENTKDIRL"
     3'UTR           1380..2017

BASE COUNT    388 a    616 c    596 g    425 t
ORIGIN
```

Query Match   Best Local Similarity 26.5%; Conservative 65.4%; Pred. No. 7.6e-85; DB 11; Length 2025;
Matches 574; Conservative 0: Mismatches 298; Indels 6; Gaps 1;   Score 379.2;

```
Qy    243 TGAGGTGGAGCTGGAGCTGGTTGGTTGATGAGGATTTCAAGTTCATCCTGCTGCTTGTGAG 302
          |||| | ||                |||  | |||||||||||||||||||||||| ||||
Db    302 TGAGCTGGGCTACAGGTGCCGCTTCAACGAGGACTTCAAGTACGTGCTGCTGCCTGTC 361

Qy    303 CTATGCCAGTTGTCTTTTGTGTCTGGGCTTGGGCCTTAACGCCCCAACCCTATGGCTGTCTTCAT 362
          ||    ||  | |  ||||| | ||| ||   |||  |  ||||  |||||| |  | ||  |
Db    362
```

RESULT 9

```
LOCUS       S74902                  2025 bp    mRNA            PRI       09-MAY-1995
DEFINITION  H2PU-P2U nucleotide receptor [clone 5a2116] [human, airway
            epithelium, mRNA, 2025 nt].
ACCESSION   S74902
NID         g802034
VERSION     S74902.1  GI:802034
KEYWORDS    .
SOURCE      human airway epithelium.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2025)
  AUTHORS   Parr,C.E., Sullivan,D.M., Paradiso,A.M., Lazarowski,E.R.,
            Burch,L.H., Olsen,J.C., Erb,L., Weisman,G.A., Boucher,R.C. and
            Turner,J.T.
  TITLE     Cloning and expression of a human P2U nucleotide receptor, a target
            for cystic fibrosis pharmacotherapy
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (26), 13067 (1994)
  MEDLINE   95108098
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 158919] from the original journal article.
            This sequence comes from Fig. 1.
FEATURES             Location/Qualifiers
     source          1..2025
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            246..1379
                     /gene="HP2U"
                     /note="P2U nucleotide receptor; *ERRATUM*Proceedings of
                     the National Academy of Sciences of the United States of
                     America,Vol.91,No.8, Parr et.al.,Apr. 1994. p. 3275-3279"
     CDS             246..1379
                     /gene="HP2U"
                     /note="This sequence comes from Fig. 1. Protein sequence
                     is in conflict with the conceptual translation:
                     mismatch(359[S->F])"
                     /codon_start=1
                     /product="P2U nucleotide receptor"
                     /protein_id="AAB32966.1"
                     /db_xref="PID:g802035"
                     /db_xref="GI:802035"
                     /translation="MAADLGPWNDTINGTWDGDELGYRCRFNEDFKYVLLPVSYGVVC
                     VLGLCINAVALYIFLCRLKTWNASTTVFMFHLAVSDALYAASLPLLVYYARGDHWPFS
                     TVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPLRSLRWGRARYARVAGAVWVL
                     VLACQAPVLYFVTTSARGGRVTCHDTSAPELFSRFVAYSSVMLGLLFAVPFAVILVCY
                     VLMARRLLKPAYGTSGGLPRAKRKSVRTIAVLVAVFALCFLPFHVTRTLYLYSFRSDL
                     SCHTLNAINMAYKVTRPLASANSCLDPVLYFLAGQRLVRFARDAKPPTGPSPATPAAR
                     RLGLRRSDRTDMQRIGDVLGSSEDSRTESTPAGSENTKDIRL"
BASE COUNT      388 a    616 c    596 g    425 t
ORIGIN
```

Standard Confusion A

```
Query Match          26.58%;  Score 379.2;  DB 11;  Length 2025;
Best Local Similarity 65.4%;
Matches  574;  Conservative   0;  Pred. No. 7.6e-85;
           Mismatches  298;  Indels   6;  Gaps   1;
```

Qy   243 TGAGGTGGAGCTGGACTGTTGGTTTGATGAGGAGATTTCAAGTTCATCCTGCTGTGAG 302
           ||||||| || | || ||||||| || || |||||| ||| |||| | |||
Db   302 TGAGCTGGGCTACAGGTGCCGCTTCAACGAGGACTTCAAGTACGTGCTGCTGCCTGTGTC 361

Qy   303 CTATGCAGTTGTCTTTGTGCTCGGGCTGGCTTGGGCCTTAACGCCCAAACCTATGGCTCAT 362
           ||| ||||||||| | ||| || || ||||| || || ||| || ||| |||| |||
Db   362 CTACGGCGTGGTGTGCGTGCTTGGGCTGTGTCTGAACGCCGTGGCTCTACACTTCTT 421

Qy   363 CTTCCGCGCTCCGACCCTGGGATGCAAGCGGCCACCTACATGTTCCACCTGGCATTGTCAGA 422
           | ||||| || | | ||||| || | || |||||| ||| ||| |||||| |||| |||
Db   422 GTGCCGCCTCAAGACCTGGAATGCGTCCACCACCGTATTCATGTTCCACCTGGCTGTGTCAGA 481

Qy   423 CACCTTGTATGTGCGTGCTGCTCGCCCACCCTCATCTACTTACTTATGCAGCCCACAACCACTG 482
           | |||| |||| || || | | || ||||| ||| || | ||| ||||| ||| ||| ||||
Db   482 TGCACTGTATGCGGCCTCCCTGCCGCTGCTCGTCTATTACTACGCCCGCGGGGACCACTG 541

Qy   483 GCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGGCTTCTTTTCTATTATGGAACCTCTACTG 542
           |||||| ||| ||||| |||| ||| |||| ||||| ||||| | |||||| ||||| ||||
Db   542 GCCCTTCAGCACGGTGCTCTGCAAGCTGGTGCGGTTCCTCTTCTACACCAACCTTTACTG 601

Qy   543 CAGTGTCCTTTTCCTCACCTGATCACGGTGACACGGCTACCTGCCACCCACT 602
           ||| |||| ||| |||| || |||| | |||| || ||| || ||| ||||
Db   602 CAGCATCCTCTTCCTCACCTGCATCAGCGTGCACCGGTGTCTGGGCGTCCTGAGACCTCT 661

Qy   603 TCGGGCCACTACGCCTGGGGGCGCCGGCTGCCAGGCCTTCTCTGCCTGCCAGTTTGGTT 662
           || | ||| ||| || ||| || | ||| ||| || ||| ||| ||| || |||||
Db   662 GCGCTCCCTGCGCTGGGGCCGCGCCCGCTACGCCCGGGTGGCTGGGGCTGTGTGGGT 721

Qy   663 GGTCGTAGCCGGCTGCCGTGGCCCAACCTGTTCTTTGTGTCACAACACCAGCACAAAGGGAC 722
           || | | || |||| || ||| ||| || | ||| | ||| || || |||| |||
Db   722 GTTGGGTGCTGGCTGCCGCCAGGCCCCGGTGCTCTACTTTGTCACCACCAGCGCGCGGGGG 781

Qy   723 CACCGTCCTGTGCCCATGAGAACCACTCGGCTGCCAGAGAGTTTGACCACTATTGTCACTTCAG 782
Db   782 CCGCTAACGTGCCACGACACCTCGGCACCCGAGCTCTTCAGCCGCTTCGTGGCCTACAG 841

Qy   783 CTCGGGGGTCATGGGGCTGCTCTTTGCGGTGCCCTTCGCTGTGATCCTGGTCTGTTATGG 842
Db   842 CTCAGTCATGCTGGGGCTGCTCTTTGCCGTGCCGTTCGCTGTCATCCTTGTGTGCTACGT 901

Qy   843 ACTCATGGCCGTCGTCGCTGTCGTGTGATACAGCC------CTTGCCCAGGCTGCTGCCACAGTCGTCGTTC 896
Db   902 GCTCATGGCCCGGCGACTGCTAAAGCCAGCTTACGGGACCTCCGGAGGCCTCCCTAGGGC 961

Qy   897 TCGCCTTCCGCTCTCCGGCACCATAGCCTGTGGTGGTCGACTGTCTTTTGCTGTCGTTCGT 956
Db   962 CAAGCGCAAGTCCGTGCGCACCATCGCCGTGCTGGTGGCTGTCTTCGCCCTGTGCTTCCT 1021

Qy   957 GCCTTTCCACATCACCCGCACCATTTACTACCTGTGGCCAGGCGTTGGTGAAGCTGACTGCCG 1016
Db   1022 GCCATTCCAGGTCACCACCGACCCTACTACTCCTTCCGCTCGGACCTCTCATGGCTCTCAT 1081

Qy   1017 AGTACTGAACACATTGTCAACGTGGTCTATAAAGTGACTCGGCCCCTGGCCAGTGCCAACAG 1076
Db   1082 CACCCTCAACGCCATCAACATGGCCTACAAGGTTACCCGGCCGCTGGCCAGTGCTAACAG 1141

Qy   1077 CTGCCTGGGATCCTGTGTGCTCTACTTGCTCACTGGGGGGACA 1114
Db   1142 TTGCCTTGACCCCGTGCTCTACTTCCTGGCTGGGCAGA 1179

Sequence Comparison A (cont.)

```
RESULT    10
RATP2U
LOCUS       RATP2U      2011 bp    mRNA        ROD     07-JUN-1996
DEFINITION  Rattus norvegicus G protein-coupled purinoceptor (P2U) mRNA,
            complete cds.
ACCESSION   L46865
NID         g960286
VERSION     L46865.1  GI:960286
KEYWORDS    .
SOURCE      Rattus norvegicus (strain Wistar) (clone: rpP2U) adult pituitary
            gland cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Chen,Z.P., Krull,N., Xu,S., Levy,A. and Lightman,S.L.
  TITLE     Molecular cloning and functional characterization of a rat
            pituitary G protein-coupled adenosine triphosphate (ATP) receptor
  JOURNAL   Endocrinology 137 (5), 1833-1840 (1996)
  MEDLINE   96198783
FEATURES             Location/Qualifiers
     source          1..2011
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone="rpP2U"
                     /tissue_type="pituitary gland"
                     /dev_stage="adult"
                     /note="(vector lambda ZAPII)"
     5'UTR           1..483
     gene            484..1608
                     /gene="P2U"
     mat_peptide     484..1605
                     /gene="P2U"
                     /product="G protein-coupled purinoceptor"
     CDS             484..1608
                     /gene="P2U"
                     /codon_start=1
                     /evidence=experimental
                     /note="cloned from pituitary cells"
                     /product="G protein-coupled purinoceptor"
                     /protein_id="AAB02099.1"
                     /db_xref="PID:g960287"
                     /db_xref="GI:960287"
                     /translation="MAAGLDSWNSTINGTWEGDELGYKCRFNEDFKYVLLPVSYGVVC
                     VLGLCLNVVALYIFLCRLKTWNASTTVMFHLAVSDSLYAASLPLLVYYAQGDHWPFS
                     TVICLKLYFLFTNLYCSILFLTCISVHRCLGVLRPLHSLSWGHARYARRVAAVWVL
                     VLACQAPVLYFVTTSVRGTRITCHDTSARELFSHFVAYASSVMLGLLFAVPSIILVCY
                     VLMARRLLKPAYGTTGLPRAKRKSVRRIALVLAVFALCFLPFHVTRILYYSFRSLDLS
                     CHTLNAINMAYKIITRPLASANSCLDPVLYFLLAGQRLVRFARDAKPEPTPSPQARRK
                     LGLHRNRTDDYVRKDLSISSDDSRRTESTPAGSETKDIRL"
     3'UTR           1609..2011
     polyA_site      1609
                     2011
BASE COUNT      398 a    615 c    570 g    428 t
ORIGIN

Query Match
Best Local Similarity  25.1%;  Score 359.2;  DB 12;  Length 2011;
Matches  560;  Conservative  63.9%;  Pred. No. 7.7e-80;  Indels 3;  Gaps 1;
     560: ; 0;  Mismatches 313;
```

```
Qy  244  GAGGTGGAGACCTGGACTGTTGGGTTTGATGAGGATTTCAAGTTCACGACCACACAGCTGG  303
         |||  |||     ||  ||  | | | |||  || |||||| |||||||| |  ||||  ||
Db  541  GAACTGGGTTACAAAATGTCGCTTCAACGAGGACTTCAAGTATGTGCTGCCCGTGTCC     600

Qy  304  TATGCAGTTGTCTTTTGTGCTGGGCTTGGGCCTTAACGCCCCAAACCCTATGGCTCTTCATC  363
         ||||  | ||| ||  | | ||||| | ||||| |||  | ||||||  |  ||| ||||| |
Db  601  TATGGCGTGGGTGTGCGGTCCTCGGGCTGTGCCTGAAACGTCGTGGCCCTACACTTCCTG   660

Qy  364  TTCCGCCTCCGACCCTGGGATGCAAACGGCCTCACCACCTACACAGAGTCCACACTGGCATTGCAGAC  423
         |||||||| || ||||||||||  | |||||  ||||| ||||| ||| ||| |||||  ||||||||
Db  661  TGCCGGCCTCAAGACCTGGGAAACGGCCTCCACACTAGTTCACCTGCCCAGTTCTGAC       720

Qy  424  ACCTTGTATGTGCTGTGCCTGCCCACCCTCATCTACTATTATGCAGCCACCAACCACTGG  483
         ||| | | |||||||| |||| ||  || |||| || |  ||| || ||| |||||||||
Db  721  TCTCTCTAACGCAGCCCTCCCTGCCGCGTCGGTTTATTACTACCCCCAGGGTGACCACTGG   780

Qy  484  CCCTTTGGCACTCGAGAATCTGCAAGTTCGTGCCGCGTTTCTTTTCTTCTTAATGGAACCTCTACTGC  543
         ||  | |||| | |||||| || ||  | || |||| ||  || |||| ||| |  || |||||||||
Db  781  CCAATTACCACAGTGCTCTGCAAGCTGTGCGTTCCTTTTCTTTCTTAACACTCTACTGC      840

Qy  544  AGTGTCGTTTTCCTCACCTGCATCAGCGGTGCACCGGCACGGCCATCTGGCACTT  603
         || | | ||| ||||||||||||||| | ||| |||| | |||||| || || ||
Db  841  AGCATCCTTCTTCCTCACCTGCATCAGCGTGCACCGGTGCCTGGGCGTCCTGCGCCCTG    900

Qy  604  CGGGCACTACCGCTCGGGGCGCCCCTCGCCGCGAGGCCCTTCTCTCGGCCCGGCAGTTTGGTTG  663
         ||||||||| ||||| ||||| || || || || ||||| |||| ||| |||| ||||| | ||
Db  901  CACTCCCTGAGCTGGGGCCATCGCCGCTATGCCCGCGTGCGCGTTGTGTGGGGTG          960

Qy  664  GTGCTAGCCCGGGCTGCCCTGCGTGCCCAACCTGTTCTTTGTGCACAACCAGCAAACAAAGGGGAC  723
         ||| | || ||||| |||| ||||| |  || |||||||||||||| || ||| |||||||||||
Db  961  CTPAGTGCTGGCCGGCCGGTGCTCTACTTTGTCACCACCAGCGGTGAGAGGGGACC        1020

Qy  724  ACCGTCTGTGCCATGACAACTCGGCCTGAAGAGTTTGACCACTAYGTGCACTCAGC  783
         ||| |||||||||||||||||||||||||| |||| |||||||| |||||| ||||
Db 1021  CGAATCACCTGCCACGACACACCTCGGAGCTCTTAGCCCATTTTGTGGCTTACAAC        1080

Qy  784  TCGGCGGTCATGGGGCTGCTGTTTGGGTGCTGTTTGGCTGCTGTCGTTGCTTATGGA  843
         |||||||||||||||||||||||| ||| |||| || ||| || ||||||  ||| |
Db 1081  TCTGTCATGCTGGGTCTGCTTTTTGCTGTGCCCTTTCCCATCANTCCTGGTCGTTACGTG    1140

Qy  844  CTCATGGCTCGTGCCGTCTGTATCAGCCCTCTGCACAGTCGTCTTCTGCCCTC  903
         ||||||||  ||||| ||||| || |||  |  ||||| ||| ||||||| |
Db 1141  CTCATGGCCCGACGGCTGCTCAAAACCGGCTTATGGGACCACAGGTCTGCCCTCGGGCAAG   1200

Qy  904  CGG---TCTCTCCGACCATAGCCTGTGGTGCTGACTGTCTTGCTGTTCGTCGT  960
         |||     | |||| |||| |||  |||||||||| ||| ||| | ||| |||
Db 1201  CGGAAGTCTGTGCGCCACCATCGCCCTTAGTACTGGGCCGTCTTCGCCCTCTGCTTCCTG    1260

Qy  961  TTCCACANCACCCGCACCCGCCACCATTTACTTACCTGCCGGTGTTGGAAGCTGAGTCCGGAGTA 1020
         |||||| ||||||||||| || ||||| ||| |||| || | |||||||  |||||   || |
Db 1261  TTCCACGTCACCCGCACCCTCTATTACTCCTTCCGATCACTCCGCAGTCCTCCACACC       1320

Qy 1021  CTGAACATTGTCAACGTGGGTCTATAAAGTGACTCGGCCCCTGGGCTGCCAACAGCTGC 1080
         || |||  || ||| | ||| |||||| | || |||||  |||||||| |||||| ||||
Db 1321  CTCAACGCCATCAACATGGCGTATAAGATCACCCGGCCACTGGCCAGCGCCAACAGTTGC    1380
```

```
Qy  723  CACCGICCTGTGCCATGACAACTCGGCCTGAAGAGTTGTGGACCARTGTGCACTTCAG  782
         |||    || | || |||||| | ||||  |||||| | || |||  |||| ||| |||
Db  782  CCGCGTAACCTGCCACAGACACCTCGGCACCGAGCTCTTCAGCACTAGTGTGGCCTACAG   841

Qy  783  CTCGGCGGGTCATGGGGCTGCTGTTTGCGTGCCGCTGCCTGGTCACTCHTGTTAATGG   842
         ||| | | |||||||||||||| ||| ||||| | ||  | ||| |  | || |  | |
Db  842  CTCAGTGATGTGGGCCTGCGTCCTTTCGGCGTGCCCTTTGCGTCATCCTTTGTCGTTACGT   901

Qy  843  ACTCATGGCTCGTGCTGCCTGTANCAGCC----CTTGCCAGGCTCTCGCACGTCCTTC   896
         ||||||||||||  | |||| | | |||||    || ||||| ||||| ||| |||
Db  902  GCTCATGGCTCGGCGACTGCTAAAGCCAGCCTTARAAGCCAGCCGCCTCCCTAGGGC      961

Qy  897  TCGCGTCCTCGCTCTCCGCACCCATAGCTGTGTGTGTGCTGTCTTTGCTTGCTCGT   956
         ||| ||||| ||| ||||||| |||||| ||| |||| ||| |||||| ||| |  ||
Db  962  CAAGCGCAAGTCGCTGTCGCCACCATCGCCCGTGGCTGCTGTCTTCCGCCCCGCTTCCT   1021

Qy  957  GCCTTTCACAACACCCGCACCCGCACCCATTTTACTACCTGGCCAGGCGCGTGTGGAAGCTGACTGGC 1016
         |||  ||||| ||||| ||||| |||    ||| |||| | | | ||  || |  || |||||  ||
Db 1022  GCCAATTCCACGTCACCCGACCCCTCTACTACTTCGTGGACTGAGCCTCACGGTCCCA        1081

Qy 1017  AGTACTGGAACATTGTCAACGGTGGTCTTATAAAGTGACTGGGCCCCCGGCCC      1076
         || |  | |||| ||| ||| | || || ||||| | |||| |  || ||  |
Db 1082  CACCCTCAACCGCCATCAACACATGGCCTACAAGGTTACCCGCGCCGGCGCCTAACAG      1141

Qy 1077  CTGCCTGGATCCTGTCTCTTACTTGCTCACTGGGGACA     1114
         ||||| | ||| |||| || ||| | || | ||| ||||
Db 1142  TTGCCTTGACCCGTGCTCTCACTTCCTGGCTGGGCAGA    1179
```

```
Qy    1081  CTGGATCCTGTGTCTACTTGCTCACTTGCTCACTGGGACAAA 1116
            || |||||| |||||||||| || ||| |||| | ||| |||
Db    1381  CTTGACCCTGTGCTCTACTTCCCTGGCCAGGGCAGAGA 1416
```

RESULT 11

```
Db    1381  RNU56839
```

LOCUS   RNU56839   1688 bp   mRNA   ROD   03-FEB-1998
DEFINITION Rattus norvegicus P2u receptor protein mRNA, complete cds.
ACCESSION   U56839
NID   g1336124
VERSION   U56839.1 GI:1336124
KEYWORDS   .
SOURCE   Norway rat strain-Wistar Kyoto.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 1688)
AUTHORS   Seye,C.I., Gadeau,A.P., Daret,D., Dupuch,F., Alzieu,P., Capron,L.
            and Desgranges,C.
TITLE   Overexpression of P2Y2 purinoceptor in intimal lesions of the rat
            aorta
JOURNAL   Arterioscler. Thromb. Vasc. Biol. 17 (12), 3602-3610 (1997)
MEDLINE   98099857
REFERENCE   2 (bases 1 to 1688)
AUTHORS   Seye,C.I., Gadeau,A.P. and Desgranges,C.
TITLE   Direct Submission
JOURNAL   Submitted (26-APR-1996) Claude Desgranges, U 441, INSERM, avenue du
            Haut-lavaque, Pessac 33600, France
FEATURES   Location/Qualifiers
     source   1..1688
                     /organism="Rattus norvegicus"
                     /strain="Wistar Kyoto"
                     /db_xref="taxon:10116"
     CDS   142..1266
                     /note="receptor"
                     /codon_start=1
                     /product="P2u"
                     /protein_id="AAC00048.1"
                     /db_xref="PID:g1336125"
                     /db_xref="GI:1336125"
                     /translation="MAAGLDSWNSTINGTWEGDELGYKCRFNEDFKYVLLPVSYGVVC
                     VLGLCLNVVALYIFLCRLKTWNASTTYMFHLAVSDSLYAASLPLLVIYYAQGDHWPFS
                     TVLCKLVRFLFYINLYCSILFLTCISVHRCLGVLRPLHSLSWGHARYARRVAAVWVVL
                     VLACQAPVLYFVTTSVRGTRITCHDTSARELSHFVAXSSVMLGLLFAVPFSILLVCY
                     VLMARRLLKPAYGTVGLPPRAKRKSVRTIALVLAVFALCFLPFHVTRTLYYSFRSLDLS
                     CHTLNAINMAXKITRPLASANSCLDPVLYFLLAGQRLVRFARDAKPATEPTPSPQARRK
                     LGLHRPNRTDVRKDLSISSDSRRTESTPAGSETKDIRL"
ORIGIN
BASE COUNT   357 a   513 c   443 g   375 t
```

```
Qy     244  GAGGTGGAGCTGGACTGGGTTGGTTTGATGAGGATTTCAAGTTCATCATCCTGCTGCCTGTGAGC 303
            ||    ||  |           ||||| |||| |  ||||||||  ||| | ||  | ||||
Db     199  GAACTGGGTTACAAAATGTCCGCTTCAAGAGGACTTCAAGTATGTGCTGCTGCCGGTGTCC 258

Qy     304  TATGCAGTTGTGTCTTTGTGCCTGGGCCTTGGGCCTTAACGCCCAACCTATGGGCTTCTCATC 363
            ||||  | |||| ||||| |||| ||||  |    ||||   ||||| ||| || ||| |||
Db     259  TATGGCGTGGTGTGCGTGCTCGGGCTGTGCCTGAACGTCGTGGCCCTCTACATCTTCCTG 318

Qy     364  TTCCGGCTCCGACCCTGGGATGCGCCACCTACACTTGCCACCTGGCATGTGCAGAC 423
            ||| |   | || |||| || | |||  || ||| || ||  ||||  || || |
Db     319  TGCGGCCTCAAGACCTGGAACGCCTCCACCACCTACATGTTCCACTTGGCAGTTGATC 378

Qy     424  ACCTTGTATGTGCTGCTGCCTGCCCACCCTCATCTACTACTTATGCCAGCCCACAACCACTGG 483
            ||    | ||| |  | ||| ||||  | || | ||| || | | | || || ||||||||
Db     379  TCTGTCTACGGAGCCTCCCTGCCGCTGCTGGTTATTTATTACACGCCCAGGGTGACCACTGG 438

Qy     484  CCCTTTGCACTCAAGATCTGCAAGTTCGTCCGGCTTCGTTCTTTTCTTATTGGAACCTCTACTGC 543
            ||  || ||||| || || ||||| || ||| ||| ||| ||    | |||  | || |||||
Db     439  CCATTTAGCACAGTGCTCTGCAAAGCTGTCTCGAAAGCTGGTGCGTTTCCTTTTTACACTAGCTGC 498
```

Query Match   25.1%; Score 359.2; DB 12; Length 1688;
Best Local Similarity   63.9%;
Matches 560; Conservative 0; Mismatches 313; Indels 3; Gaps 1;

RESULT 12

```
Db    1039  RNU09402
```

LOCUS   RNU09402   1918 bp   mRNA   ROD   26-JAN-1995
DEFINITION Rattus norvegicus purinergic receptor (P2u) mRNA, complete cds.
ACCESSION   U09402
NID   g563807
VERSION   U09402.1 GI:563807
KEYWORDS   .
SOURCE   Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 1918)
AUTHORS   Rice,W.R., Burton,F.M. and Fledelley,D.T.
TITLE   Cloning and expression of the alveolar type II cell P2u purinergic
            receptor
JOURNAL   Am. J. Respir. Cell Molec. Biol. 12, 27-32 (1995)
REFERENCE   2 (bases 1 to 1918)
AUTHORS   Rice,W.R.
TITLE   Direct Submission
JOURNAL   Submitted (05-MAY-1994) Ward R. Rice, Div. of Pulmonary Biology,
            Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati,
            OH 45229-3039, USA
FEATURES   Location/Qualifiers
     source   1..1918
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /cell_type="alveolar type II cell"
     5'UTR   1..358

```
Qy     439  GTCGTTAGCCCGGCTGCCTCGTGCCCAAACCTGTTCTTGTTGTCACAACCAGCAACAACAAAGGACC 498
            || ||     ||  | | | ||| ||| | |||| | |||||| | || ||| ||| |||
Db     499  AGCACCCTTCTTCCTCACCCTGCATCAGCGTGTGCCACCGTGCCTGGGGGTCTTCGCCGTCCCACT 558

Qy     499  CTAGTGACCTGGACGGCCTGCCAGGCGACCCCGTGGCCTACTTCGTCACCACCAGCGTGAGGAGGACC 558
            || || ||||||| ||  || || ||  |||  ||||  ||| |||| || || |||| |||
Db     559  CACTCCCTGAGCTGGGGCCATGCCCGGCTATGCCCGCCGTGTGGCCGTGCCTGGGGTCCT 618

Qy     559  ACCGTTCCTGTGCCATGACACCACCTCGGGTGCCTGAAGAGTTGACCACTAGTGCACTTCAGC 618
            | || |  |||||| |  || || || ||  | || | ||||| |||| ||  | |||||
Db     619  CGAACCACCGTGCCCAGGACACCTCCGCCCTGCGGAGCTGCTTTAGCCACTTTGTGGCCTTACAGC 678

Qy     619  CGGGGCACTACGGTGGGGGCGCCCCCTGCCCAGGCGACCCCGTGCCTTGAGAGAGGTTGACCACTG 678
            | |||  |||    | ||| |||| ||||   |||| |  || ||| || || | |||
Db     679  TCGTGTCGGCCCGACCGGCTGCTCAAACCGCCCTTATTGGACTCACACTCGTTTTCTGGTTG 738

Qy     679  TCGGGCGGGGTCAATGGGGGCTGCTCTCTTTGGGGCGTGCCTCCGCTCCCCTGGCTGTATGGA 738
            || ||  ||| |  ||||| | || | ||| || |||| ||| ||| |||||| || |||||
Db     739  TCTGTGATGCTGGGTCTGCTTTTTGCTGTGCCCTTCTCCATCCTCCTGGTGTGCTACGTG 798

Qy     739  TCGGGGGTCATGGGGCTGCTCTTTGGGGCGTGCCCTCCGCTCCCCTGGCTGTATGGA 783
            |||| | |||   |  |  || | |||| | || |  ||||  ||| |||| ||||
Db     799  CTCATGGCCCGACGGCTGCTCAAACCCGCTTATGGGACAGTGGGACTGCCGCCAAG 858

Qy     784  TCGGGGGTCATGGGGCTGCTCTTTGGGGCGTGCCTCCGCTCCCCTGGCTGTATGGA 843
            | || ||  | ||| || || |  ||||  | || |||| |||||| |||| ||||
Db     859  CGCAAGTCGTGTGGCGACCATCGGCCTTAGTACTGCCCTGGTGCTGGCTGTGTTTGCCCT 918

Qy     844  CTCATGGCGCTCGTCGCCCTGTATCAGCCCTTCGCACAGTCGTCTTCCTGCCTC 903
            |||||| |||| | | ||||| |  ||| | | |||| || || | ||| |
Db     919  TTCCACGTCACCCGCACCCTCTATTACTCCTTCCGATCACTTGACCTCAGTTGCCACACC 978

Qy     904  CGC---TCTCTCCGCACCATAGCTGTGGTGGCTGACGTCGTCTTTGCTGTTCGTGTGCCT 960
            |||   |  | || ||  || || | || | || | | ||||| | ||| |||  ||||
Db     979  CTCAACGCCATCAACATGGCGTAAAGATCACCCGGCCACTGGCCAGCGCCAACAGTTGC 1038

Qy     961  TTCCACATCACCCGCAACCCACCCAGTTACTTACTTACCTGGCCAGGCTGTTGGGAAGCTGACTGCGCAGAGTA 1020
            |||||| | | ||  |||  ||  | || | |  || |||| ||  |||| | ||  || || || |||||||
Db    1019  TTCCACGTCACCCGCACCCTCTATTACTCCTTCCGATCACTTGACCTCAGTTGCCACACC 978

Qy    1021  CTGAACATTGTCAACGTGGGTCTATAAAGTGACTCGGACCCCTGCCGCCAGTTGCCCAACAGCTGC 1080
            || || |    |||||| | |  | |  | |  | | |  ||| ||  ||| |||  |||||||
Db     979  CTCAACGCCATCAACATGGCGTAAAGATCACCCGGCCACTGGCCAGCGCCAACAGTTGC 1038

Qy    1081  CTGGATCCTGTGTCTACTTGCTCACTTGCTCACTGGGACAAA 1116
            || |||||| |||||||| || |||  || | ||| |||
Db    1039  CTTGACCCTGTGCTCTACTTCCCTGGCCAGGGCAGAGA 1074
```

```
CDS            359..1483
               /gene="P2u"
               /codon_start=1
               /product="purinergic receptor"
               /protein_id="AAA61565.1"
               /db_xref="PID:g563808"
               /db_xref="GI:563808"
               /translation="MAAGLDSWNSTINGTWEGDELGYKCRENEDFKXVLLPVSYGVVC
               VIGLCLNVVALYIFLCRLKTWNASTTYMHLAVSDSLIAASLPLLVYIYAQGDHWPFS
               TVLCKLVRFLFYTNLYCSILFLTCISVHRSLGVLRPLHSLRWGHARYARRVAAVVWVL
               VLACQTPVLYFVTTSVRGTRITCHDTSDRELFSHFVAXSSVMLGLLFAVPFSILLVCY
               VLMARRLLKPAYGTTGLPRAKRKSVRTIALVLAVFALCFLPFHVTRTLYYISFRSLDLS
               CHTLNAINMAKITRPLASANSCLDPVLYFLLAGQRLVRFEARDAKPATEPTSPQARRK
               LGHRPNRTDTVRKDLSISSDDSRRIESTPAGSETKDIRL"
     gene      359..1483
               /gene="P2u"
     3'UTR     359..1483
               /gene="P2u"
               /product="purinergic receptor"
     mat_peptide 359..1480
               /gene="P2u"
               /product="purinergic receptor"

BASE COUNT    404 a   581 c   512 g   421 t
ORIGIN

Query Match   25.0%;  Score 357.6;  DB 12;  Length 1918;
Best Local Similarity   63.8%;
Matches  559;  Conservative   0;  Mismatches  314;  Indels    3;  Gaps    1;

Qy  244  GAGGTGGAGCTGGACTGGTTGGTTTGATCAAGGAGGATTTCAAGTTCATCATCCTGCCTGTGAGC  303
              |  |   ||     |||   |||  ||||| || | | ||     ||| |  |  |||||||
Db  416  GAACTGGGTTACAAAATGTCGCTTCAAGAGGACTTCAAGTAATGTCGCTGCCCGTGTCC  475

Qy  304  TATGCAGTTGTCTTTGTGTCTGGGCTTGGGCCTTAACGCCCCAACCCTATGGCTGTCTCATC  363
              ||   | |||   ||||  || |  | || |  |  ||  | |||  || | |
Db  476  TATGGCGTGGGTGGTGCGTGGCTGTGCCTGAAACGTCGTGCCCCTCTCTACACTTCCTG  535

Qy  364  TTCCGGCCTCTCCGACCCTGGGATGCAAACGGCCCACCTACTACTIGGCATTGTCAGAC  423
              | ||||||   || | |||| ||   ||| |||| ||   |  || |||||  |  ||
Db  536  TGCCGGCCTCAAGACCTGGAACGGCTCCACCACTACCAAGTTTCACCTGCAGTTTCTGAC  595

Qy  424  ACCTTGTATGTCGCTGCCCTGCCCACCCTCATCTATTATCCAGCCCACAACCACTGG  483
              | | | ||  | || | | || ||| |  |  |||  |||  |||  |||| ||
Db  596  TCTCTCTACGCAGCCTCCCTGCCTGGTTTATTATTACTACGCCCAGGTGACCACTGG  655

Qy  484  CCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTTCTGTTCTATTTTCTTTTTTGGAACCTCTACTGC  543
              |||| ||| || |   ||| ||| |||  | | | | |   | || || | ||||||||||
Db  596  CCATTTAGCACAGTGCTCTGCAAGCTGGTGCGTTTCTTTTTCTTTACTAACCTCTACTGC  655

Qy  544  AGTGTGCTTTTTCCTCACCTGACATCAGGGTGCACGGGCATCGCACCACCACTT  603
              | | ||  ||  | |  | | | |||||| || |||| || ||||| |||||
Db  656  AGCATCCTCTTCCTCACCTGCATCAGCGTCCACCGGAGCCTGGGGGTCCTGCGCCCTCTG  715

Qy  604  CGGGCACTACGGTGGGGGCCCCCGCGAGGCCTTCTCTGCCTGGCAGTTTGGTTG  663
              |||| |  |  |||||  | |  | ||  |  || | |   | |||| ||| ||
Db  716  AGCATCCTCTTCCTGACATCAGCGTCCACCGGAGCCTGGGGGTCCTGCGCCCTCTG  775

Qy  664  GTCGTAGCGCGGCGTGCCTCGTGCCCAACCTGTTTCTTTGTCACAACAAAAGGACC  723
              || | |  ||  |||||  | ||  |   || |    |  ||||  | |  ||||
Db  776  CACTCCCTGCGGTGGGGCCATGCGCCCTTATGCCGCCACCGGTGAGAGGGACC  835

Qy  724  ACCGTCTGTGCCATGACACACTCCGGCCTGAAGAGTTTGACCACTCTACTG  783
              ||||| ||||  |||||| ||||| |  | |  | |   |  ||| |||
Db  836  CTAGTGCCTTGGCGTGCCCAGACAACCGGTCGCTCTCACTTTTGTGCTTACGTG  895

Qy  784  TCGGGCGGTCATGGGGCTGCCGCCCTGGCTGCCTGCCTGGAACTTCTACTG  843
              ||| |  | || | | || ||||| | |||||| |  |  |  |||| ||
Db  896  CGAATCACCTGCCAGACAACCTTTAGCCATTTGTTGTGGCTTACAGC  955

Qy  844  CTTCATAGGCTTCGTCGCCCTGTATCAGCCCTTGCCAGGCTCTGCACAGTGGTCTTCTCGCCTC  903
              || | ||| | ||| ||| |  ||   | | |||  |  || |||||  ||| | | |||
Db  956  TCTGTCATGCGTGGGTCTGCGTTTTTGCTGGTGCCTTTTTCTTTACTGTG  1015

Qy  904  CGC---TCTCTCGCACCATAGCTGTGTGGTGCTGACTGTCTTGCGTCTGCTTCGTGCGCCT  960
              |   ||||| | || || ||  | || || | ||| ||   ||| | || | | ||
Db  1016 CTCATAGCCCGACGGGTGCTCAAAACCGGCTTATGGGACCACAGGTCGTGCCTCGGGCCAAG  1075
```

```
RESULT 13
LOCUS       I34573       1842 bp  DNA             PAT
DEFINITION  Sequence 1 from patent US 5596088.
ACCESSION   I34573
NID         g1825364
VERSION     I34573.1  GI:1825364
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1842)
  AUTHORS   Boucher,R.C., Weisman,G.A., Turner,J.T., Harden,T.K., Parr,C.E.,
            Sullivan,D.M., Erb,L.J. and Lustig,K.D.
  TITLE     DNA Encoding the human P.sub.2U receptor and null cells expressing
            P.sub.2U receptors
  JOURNAL   Patent: US 5596088-A 1 21-JAN-1997;
FEATURES             Location/Qualifiers
     source          1..1842
                     /organism="unknown"

BASE COUNT    364 a   558 c   527 g   393 t
ORIGIN

Query Match   24.6%;  Score 352;  DB 5;  Length 1842;
Best Local Similarity   65.0%;  Pred. No. 4.8e-78;
Matches  571;  Conservative   0;  Mismatches  295;  Indels   12;  Gaps    3;

Qy  243  TGAGGTGGAGCTGGACTGGTTGGTTTGATGAGGAGGATTTCAAGTTCATCCTGCCTGTGAGC  302
              |||| |||| |  |  |||| |||||| |||||| ||||||||||  ||| |||||||||
Db  113  TGAGCTGGGCGCTACAGGGTGCCGCGCTTCAACGAGGACTTCAAGTACGTGCGTGCTGTGTC  172

Qy  303  CTATGCAGTTGTCTTTGTGTCTGGGCTTGGGCCTTAACGCCCCAACCCTATGGCTGTCTCAT  362
              || | |||  |  | ||| | || |   ||  || | || ||  ||   | |||  |  |
Db  173  CTACGGCGTTGGTGGTGCGTGGCTGTGCCTGAAACGTCGTGGGCCTTCTACACTCTCTACT  232

Qy  363  CTTCCGGCCTCTCCGACCCTGGGATGCAAGCGCCACCTACTACTCAATCTACTATTATGCAG  422
              |  ||||| ||   | ||| |  ||   | || ||||| || | ||  | |   || |||
Db  233  GTGCCGCCTCAAGAGCCTGGAATGCGTCCAAGGCCACTCACTACTACGCCCAGGTCACCACTG  292

Qy  423  CACCTTGTATGTCGCTGCCCTGCCCACCCTCATCTACTACTATTATGCGCCCGGGGACCACTG  482
              | | | | |  | || |   || || | | | | ||| ||||  || ||| |   |||| |
Db  293  TGCACCTGTATGTCGCTGCCCTCAGGCCGCTGCACCACCGGTGAGAGGGACCACAACCACTG  352

Qy  483  GCCCTTTGGCACTGAGATCTGCAAGTCGTCCGCTTTTCTGTTCTATTGGAACTTCTACTG  542
              | ||| || || | | |||||||| |  ||||  |||| | |   |   ||| ||||
Db  353  GCCCTTCAGCACGGTGCTCTGCAAGCTGGTGCGTTTCCTTCTTCTTTACTACGCCCACTG  412

Qy  543  CAGTGTCTCTTTTCCTCACCTGCATCAGCGTGCACCGGAGCCTGGGGGTCCTGCGCCCTCTG  602
              ||||| |  | | ||| | |  | |||| | |  ||| |||  ||||| ||||| |||||
Db  413  CAGCATCCTCTTCCTCACCTGCATCAGCGTCGACGGGTGTCTGGGCGTTCTGCGCCCTCTG  472

Qy  603  TCGGGCACTACGGTGGGGGCCCCCGCGAGGCCTTCTCTGCCAGGCCGTTGGTT  662
              |||| |  ||||| ||  |  |   |  |  ||| | |||||  | |  |||||
Db  473  GCGCTCCCTGCCTGCGGGGCCCCGCGCTACCGCTGCAGTTTGGTT  532

Qy  663  GGTCGTAGCCGGCGCTGCCTGCCCAACCTGTTTCTTTGTCACAACCAGCAACAAAAGGGAC  722
              || |  |||||| |  | || || || |||| |||  ||||| | || || |  || ||
Db  663  GGTCGTAGCCGGCGCTGCCTGCCCAACCTGTTTCTTTGTCACAACCAGCAACAAAAGGGAC  722
```

RESULT 14

```
LOCUS        I36524                 1842 bp    DNA                PAT
DEFINITION   Sequence 1 from patent US 5607836.
ACCESSION    I36524
NID          g2086349
VERSION      I36524.1  GI:2086349
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 1842)
  AUTHORS    Boucher,R.C., Welsman,G.A., Turner,J.T., Harden,T.K., Parr,C.E.,
             Sullivan,D.M., Erb,L.J. and Lustig,K.D.
  TITLE      Methods of detecting compounds which bind to the P.sub.2U receptor
  JOURNAL    Patent: US 5607836-A 1 04-MAR-1997;
FEATURES             Location/Qualifiers
     source          1..1842
                     /organism="unknown"
BASE COUNT      364 a    558 c    527 g    393 t
ORIGIN
```

```
Qy    533 GTTGGGTGCTGGCGCTGGCGCTGCCCAGGCGCTCTACTTGTCACCACCAGCGCGGCGGGG-- 591
              |||||||         |||||
Db    533 CACCGTCGTGTGCCATGACAACCACTGAGGCCCCCGTGCTCTACTTGTCACTTCAG 591

Qy    723 CACCGTCGTGTGCCATGACAACCACTGAGGAGTTTGACACTTCAG 782
              |||||||  |  ||||   ||||| | |||
Db    591 -CCGGTAACCTGCCGACGACACCTCGGCCGACAGAGTCTTCAGCCGCTTC 649

Qy    783 CTCGGGCGGTCATGGGGCTGCTCTCTTTGGCGTGCCGTGCCCGAGCGTCACTCTTGTTTGCTATGG 842
              |||||  |||||||||  | |  |  ||||||||  ||| |||||||
Db    650 CTCAGTCATGCTGGGCCTGTCTTCGCGTCAATCCTTGTCTGTTACGT 709

Qy    843 ACTCATGGCTGCTGCGTGCGCCTGTATCAGCC------CTTGCCAGGCTCTGCACAGTCGTCTTC 896
              |||||  ||  |  |||    |   ||  ||
Db    710 GCTCATGGCTCGGCGACTCCTAAAGCCAGCCTAGGGACGGGGGCCTGCCTAGGGC 769

Qy    897 TCGCCTCCGCTCTCTCCCACCCATAGCTGTGTGGTCGTGTCTTTGCTGTCTCGCTTCGT 956
              |||||        || | |||  |||||||||||||  | |   | ||||
Db    770 CAAGCGCAAGTCCCTGCCGCACCCATAGCTGTGGCGTCTCTTTTCCGCTATCTTTCCT 829

Qy    957 GCCTTTCCACATCACCCGCCACCATTTACTACCTGGCCAGGCGTGTTGGAAGCTGACTGCCG 1016
              ||| |   |||| | |  ||| |  |  | |  |||  ||| ||   |   || ||
Db    830 GCCATTCCACGTGCACCCCGCCACCCTCTACTACTCCCTCCCTGCCTGGCTTCGGGGCCTACAC 889

Qy   1017 AGTACTGAACATTGTCAACGTGGTTGTATAAAAGTGACTCGGCCCCCTGGCCAGTGCCAACAG 1076
              ||| |   |||| | |||   ||||| | |||||||||||| |||||||||
Db    890 CACCCTCAACGCCATCAACATGGGCTACAACATGGCCTACAGAGGTTACCCGG---CTGGCCAGTGCTAACAG 946

Qy   1077 CTGCCCTGGATCCTGTGCTCTACTTGCCTCACTGGGGACA 1114
              ||||  ||  |   ||||  ||||| |||||||||
Db    947 TTGCCTTGACCCGTGCTCTACTTCCTGCCTGCTGGGCCAGA 984
```

```
Query Match             24.6%;     Score 352;     DB 5;    Length 1842;
Best Local Similarity   65.0%;     Pred. No. 4.8e-78;
Matches   571;  Conservative   0;  Mismatches  295;  Indels  12;  Gaps   3;
```

RESULT 15

```
LOCUS        I76280                 1842 bp    DNA                PAT
DEFINITION   Sequence 1 from patent US 5691156.
ACCESSION    I76280
NID          g3012434
VERSION      I76280.1  GI:3012434
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 1842)
  AUTHORS    Boucher,R.C., Welsman,G.A., Turner,J.T., Harden,T.K., Parr,C.E.,
             Sullivan,D.M., Erb,L.J. and Lustig,K.D.
  TITLE      Method of inhibiting cell growth with the P.sub.2U receptor
  JOURNAL    Patent: US 5691156-A 1 25-NOV-1997;
FEATURES             Location/Qualifiers
     source          1..1842
                     /organism="unknown"
BASE COUNT      364 a    558 c    527 g    393 t
ORIGIN
```

```
Db    293 TGCACTGTATGCGGCCTCCCGTGCCGGTCGCGTGCGGGTCGTCGTTATTACTACGCCCGCGGGGCGACCACTG 352
Db    353 GCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTGTCTTTTCTTATTGGAACCTCTACTG 412
Qy    543 CAGTGTCCTTTTCCTCACCTTCATCAGCGTGCACCGCTACCGTGGCGCCATCTGCCACCCACT 602
Db    413 CAGCAACCCTCTTCTCCTTTCACCTCGCATCAGCCGTGCACCGGTGTCGTGGGGCGTCTTACGACCTCT 472
Qy    603 TCGGGCACTACGGCTGGGGCGCCGGCCTGCCGCCAGGCCGTTCTCTCTGCCTGCTTGGGTT 662
Db    473 GCGCGTCCCTGCCTCGGGGCGGGCGGGGCCGGCTACGCTCGGGGCCGTGTGGGT 532
Qy    663 GGTGCGTAGCCGGCGCTCGCCTTCGTGCCCAACCTGTTCTTTGTCACAACCAGCAACAAAGGGAC 722
Db    533 GTTGGTGGCTCGGGCCTCGCAGCGCCCCGTGCTCTACTTTGTCACTCAGCCGCCGGGG-- 591
Qy    723 CACCGTCCTGTGCCATGACAACCACTGGGCCTGAAGAGTTTGACCACTTCAG 782
Db    591 -CCGGTAACCTGCCGCACGAGTCTTCAGCCGCTTC 649
Qy    783 CTCGGGCGGTCATGGGGCTGCTCTCTTTGGCGTGTCGTGCCCGTGCTCAGTGCCACAGCCAACAG 842
Db    650 CTCAGTGCATGCTGGGCCTGTCTTCGCGTCAATCCTTGTCTGTTACGT 709
Qy    843 ACTCATGGCTCGTCGCGTGATCAGCC------CTTGCCAGGCTCTGCACAGTCGTCTTC 896
Db    710 GCTCATGGCTCGGGCGACTGCTAAAAGCCAGCCTAGGGACGGGGGCCTGCCTAGGGC 769
Qy    897 TCGCCTCCGCTCTCTCCGCACCATAGCTGTGTGGTCGTGACTGTCTTTGCTGTCTCGCTTCGT 956
Db    770 CAAGCGCAAGTCCCGTGCCGCACCAACGGTGGCGTCGTGTCTTCGCGCCCTTTCCT 829
Qy   1017 AGTACTGAACATTGTCAACGTGGTCTATAAAGTGACTCGGCCCCCTGGCCAGTGCCAACAG 1076
Db    890 CACCCTCAACGCCATCAACATGGCCTACAGAGGTTACCCGG---CTGGCCAGTGCTAACAG 946
Db    947 TTGCCTTGACCCCGTGCTCTACTTCCTGGCTGCTGGGCCAGA 984
```

```
Query Match             24.6%;     Score 352;     DB 5;    Length 1842;
Best Local Similarity   65.0%;     Pred. No. 4.8e-78;
Matches   571;  Conservative   0;  Mismatches  295;  Indels  12;  Gaps   3;
```

```
Qy    243  TGAGGTGGGACTGGACTGTTGGTTTGATGAGGATTTCAAGTTCATCCTGCTGCCTGTGAG  302
           ||||    |||                      |  ||||||| | |||||||||||||||
Db    113  TGAGCTGGGCTACAGGTGCCGCTTCAAACGAGGACTTCAAGTTCATCCTGCTGCCTGTGTC  172

Qy    303  CTATGCAGTTGTCTTTGTGCGTGGGGCTTGGGGGCCTTAACGCCCAACCCTATGGCTCTTCAT  362
           |||  ||| | |  | |  |||||||   ||||||||  |||  |||||||| |||  |||  |
Db    173  CTACGGCGTGGTGTTCGGTGCTTGGGGCGTGTCTGAACGCGGGCCTTCTACACTCTTCTT  232

Qy    363  CTTCCGCCCTCCGGGATGCAAACGGGCCACTCTTTACACATGTTCCACCTGGCCATTGTCAGA  422
           |||   || |                 ||||||  |  |||||||||  |  || |||||||||
Db    233  GTGCGGCCCTCAAAGACCTGGAATGCCTCCACCACATATATGTTCCACCTGGCTGTGTCTGA  292

Qy    423  CACCTTGTATGTGCTGTGCGTCGGCTGCCCACCCTTCATCTACTATTATGCCCAACAACACTG  482
           |||||||||||||  |||||||||||||||||||||||||| ||||||  |||||||||||||
Db    293  TGCACTGTATGCGGGCCTCCCTGCCGCTGCTGGTCTTATTACTACACCCCCGGGCGACCACTG  352

Qy    483  GCCCCTTTGGCACTGAGATCTGCAAGTTCGTCCCGCTTTCTTTTTCTATTGGAACCCTCTACTG  542
           ||  ||||||||||||||||||||||||  |||||||||||| |||| | |||||||||||||
Db    353  GCCCTTCAGCACGGTGCTCTGCAAGCCTGGTGGCCGCTTCCTCTTCACAACCCTTTACTG  412

Qy    543  CAGTGTCCTTTTTCCTCACCCTTGCATCAGCGTGCACCGCTACCTGGGCATCTGCCACCCACT  602
           ||  ||||||||||||| ||||||| ||||||||||||||||||||||||||||| |||||||
Db    413  CAGCAATCCTCTTCCTCACCCTGCATCAGCGTGCCACCGGTGTCTTACGACCTCT  472

Qy    603  TCGGGCACTACGGCTGGGGGGGGCCTGCCCTCGCCAGGCCTTCTCTGCCTGGCAGTTTGGTT  662
           | |       |       ||||||||||||||||||||  ||  |||||||||||||||||||
Db    473  GCGCTCCCTGCGCTGGGGGGCCGGCGGCCGCTACGGCTGCGGGGGGGGCCGTGTGGGGT  532

Qy    663  GGTGTGTAGGCCGGGCTGCCCTGGTGCCCAAACCTGTTCTTCTTTTGTCACACCAGCAACAAAGGGAC  722
           |     ||| ||||||     || ||||||||||||||||||||||||||||||||  || | |||||
Db    533  GTTGGTGCTGGGCGGCCCCGGCCAGGCCCCGGCGTGGCTCCTACTTTGTCACACCAGCGCGCGGGG--  591

Qy    723  CACGGTCTCTGTGCATGACACCACTCGCGCCTGAAGAGTTTGACCACTATGTGCACTTCAG  782
           |||||||||                ||||||||||| | ||||||||| |||||||||||||
Db    591  -CCGCTAACCTGCCACGACACCTGGCACCCGAGCTCTTCAGCGCCGCTTCGTTGGGCCTACAG  649

Qy    783  CTCGGGCGGTCATGGGGGCTGCTCCTCTTTGGGCGTGCCCCTGGCTCACTCTTGTTTGCTATGG  842
           ||| |   |||  ||||||||||||||||||| |||||||  |||| ||  |  | ||||| ||
Db    650  CTCAGTCAATGCTGGGGCTGCTCTTCGCGGGTGCCCTTTGCCGTTTCAATCCTTGTCTGTTACGT  709

Qy    843  ACTCATGGCTGCTGCCCTGTATCAGCC------CTTGCCAGGCGTCTGCACAGTGGTCTTTC  896
           || ||||||| | |  ||  || |||        |   ||| |||  |||||||||||||  |
Db    710  GCTCATGGCTCGGCGACTTCCTAAAAGCCAGCCTACGGGACCTCTGCCCCTGCCCTAGGGC  769

Qy    897  TCGCCCTCCGGCTCTCTCCGCACCATAAGCTGTGGTGGTGCTGACTGTCTTTGCTCTGCTTCGT  956
           | |     | |     | |||   || ||||||   | | ||||||||||||||||  |||||||
Db    770  CAAGCGCAAGTCCGTGCGCGCCATCGCCCTGGGTCGTTCGCCCTCTCCTTCCT  829

Qy    957  GCCTTTCCACATCACCCGCACCATTACTACCTGGCCAGGCTGTTGGAAGCTGACTGCCG  1016
           ||||||||||||||||| ||||| || |||||||||||  |||||||||  |  ||||||
Db    830  GCCATTCCACGTCACCCGCACCCTCTACTACTCCTTCCGCCTCGCCTGGACCTCAGCTGCCA  889

Qy    1017 AGTACTGAACATTGTCAACGTGGTCTATAAAAGTGACTCGGCCCCTGCCCAGTGCCACAG  1076
           ||||||||| | |  |||   |  | ||| ||   || | |||  ||| |||| |||||||
Db    890  CACCCTCAACGCCATCAACATGCCCTACAAAGGTTACCCGG--CTGGCCAGTGCTGTAACAG  946

Qy    1077 CTGCCTGGATCCTGTGTGCTCTACTTGCTCACTGGGGACA  1114
           |||   |||  | |||||||| ||||||||||||||||
Db    947  TTGCCTTGACCCCGTGCTCTACTTCCTGCCTGGGGCAGA  984
```